Truthful Learning Mechanisms for Multi–Slot Sponsored Search Auctions with Externalities

Nicola Gatti, Alessandro Lazaric, Marco Rocco, Francesco Trovò

Truthful Learning Mechanisms for Multi–Slot Sponsored Search Auctions with Externalities[☆]

Nicola Gatti[a], Alessandro Lazaric[b], Marco Rocco[a], Francesco Trovò[a]

[a] *Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milan, Italy*
[b] *INRIA Lille - Nord Europe, avenue Halley 40, 59650 Villeneuve d'Ascq, France*

Abstract

Sponsored Search Auctions (SSAs) constitute one of the most successful applications of *microeconomic mechanisms*. In mechanism design, auctions are usually designed to incentivize advertisers to bid their truthful valuations and, at the same time, to guarantee both the advertisers and the auctioneer a non–negative utility. Nonetheless, in sponsored search auctions, the Click–Through–Rates (CTRs) of the advertisers are often unknown to the auctioneer and thus standard *truthful* mechanisms cannot be directly applied and must be paired with an effective learning algorithm for the estimation of the CTRs. This introduces the critical problem of designing a learning mechanism able to estimate the CTRs at the same time as implementing a truthful mechanism with a revenue loss as small as possible compared to the mechanism that can exploit the true CTRs. Previous work showed that, when *dominant–strategy* truthfulness is adopted, in single–slot auctions the problem can be solved using suitable exploration–exploitation mechanisms able to achieve a cumulative regret (on the auctioneer's revenue) of order $\widetilde{O}(T^{\frac{2}{3}})$, where T is the number of times the auction is repeated. It is also known that, when *truthfulness in expectation* is adopted, a cumulative re-

[☆] A short, early version of this paper was presented at the ACM Conference on Electronic Commerce 2012 as: *N. Gatti, A. Lazaric, F. Trovò, A truthful learning mechanism for contextual multi–slot sponsored search auctions with externalities, in: Proceedings of the ACM Conference on Electronic Commerce (ACM EC), 2012, pp. 605–622.*

Email addresses: nicola.gatti@polimi.it (Nicola Gatti), alessandro.lazaric@inria.fr (Alessandro Lazaric), marco.rocco@polimi.it (Marco Rocco), francesco1.trovo@polimi.it (Francesco Trovò)

gret (over the social welfare) of order $\widetilde{O}(T^{\frac{1}{2}})$ can be obtained. In this paper we extend the results available in the literature to the more realistic case of multi–slot auctions. In this case, a model of the user is needed to characterize how the CTR of an ad changes as its position in the allocation changes. In particular, we adopt the *cascade model*, one of the most popular models for sponsored search auctions, and we prove a number of novel upper bounds and lower bounds on both auctioneer's revenue loss and social welfare w.r.t. to the Vickrey–Clarke–Groves (VCG) auction. Furthermore, we report numerical simulations investigating the accuracy of the bounds in predicting the dependency of the regret on the auction parameters.

Keywords: Economic paradigms, mechanism design, online learning, sponsored search auctions.

---

1. Introduction

SSAs constitute one of the most successful applications of *microeconomic mechanisms*, producing a revenue of about $6 billion dollars in the US alone in the first half of 2010 [1]. In a SSA, a number of *advertisers* bids to have their *sponsored links* (from here on *ads*) displayed in some slot alongside the search results of a keyword. SSAs currently adopt the *pay–per–click* payment scheme, which requires positive payments from an advertiser only when its ad is clicked. Given an allocation of ads over the available slots, each ad is associated with a CTR, corresponding to the probability of being clicked by the user. CTRs play a crucial role in the definition of the auction, since the auctioneer relies on (estimates of) the CTRs to determine the allocation of ads over slots and to compute the payment of each ad. Models similar to SSAs are also used in many other advertisement applications. For instance, in contextual advertising, the text of a website is scanned for keywords and an auction is used to select the ads to display in vertical/horizontal slots on the basis of the advertisers' bids and CTRs of the ads in the given context [2].

In microeconomic literature, SSAs have been formalized as a *mechanism design* problem [3], where the objective is to design an auction that incentivizes advertisers to bid their *truthful* valuations (needed for *economic stability*) and that guarantees both the advertisers and the auctioneer to have a non–negative utility. The most common SSA mechanism is the Generalized Second Price (GSP) auction [4, 5]. As shown in [4], this mechanism is not truthful and advertisers may implement bidding strategies that pay more

than bidding their truthful valuations.

While the GSP is still popular in many SSAs, the increasing evidence of its limits is strongly pushing towards the adoption of the more appealing Vickrey–Clarke–Groves (VCG) mechanism, which is already successfully employed in the related scenario of contextual advertising, by Google [2] and Facebook [6]. The first drawback of the GSP is that its equilibria may be *inefficient* (in terms of social welfare) w.r.t. the VCG outcome: considering the whole set of Nash equilibria in full information, the Price of Anarchy (PoA) of the GSP is upper bounded by about 1.6, while considering the set of Bayes–Nash equilibria the PoA is upper bounded by about 3.1 [7]. Similarly, the revenue of a (full information) Nash equilibrium can be arbitrarily small w.r.t. the VCG outcome, while in the Bayesian case the revenue is upper bounded by 6 [8]. Furthermore, the automated bidding strategies, used in practice by the advertisers to find their best bids, may not even converge to any Nash equilibrium and, under mild assumptions, the states they converge to are shown to be arbitrarily inefficient [9]. When externalities are introduced, it is known that no Nash equilibrium of the GSP provides a larger revenue than the VCG outcome [10, 11]. Finally, there is a recent increase in the use of additional features (such as larger formats, reviews, maps, or phone numbers) arranged by the search engines on the web page together with the ads to increase the attention of the user. It is known that the GSP behaves poorly in this setting, while the VCG is almost equivalent to the standard setting [12].

In this paper, we focus on the problem of designing truthful mechanisms when the CTRs are not known and need to be estimated in SSAs with multiple slots. This problem is particularly relevant in practice because the assumption that all the CTRs are known beforehand is rarely realistic. Furthermore, it also poses interesting scientific challenges since it represents one of the first examples where online learning theory and mechanism design—two important fields in artificial intelligence that recently received a lot of attention in the literature—are paired to obtain effective methods to learn under equilibrium constraints (notably the truthfulness property). For the sake of completeness, we remark that the combination of these ideas have been used also in the other fields, e.g., crowdsourcing [13].

Related works. The problem of estimating the CTRs and identifying the best allocation of ads can be effectively formalized as a *Multi–Armed Bandit (MAB) problem* [14], where each ad is an arm and the objective is to minimize the cumulative regret either on the auctioneer's revenue or the

social welfare, i.e., the difference in revenue or social welfare, respectively, of the mechanisms implemented over time estimating the CTRs and of the mechanisms that can exploit the true CTRs. The problem of budgeted advertisers (i.e., auctions where the total amount of money each advertiser is willing to pay is limited) with multiple queries is considered in [15]. This problem is formalized as a budgeted multi–bandit multi–arm problem, where each bandit corresponds to a query, and an algorithm is proposed with theoretical guarantees on auctioneer revenue regret. Nonetheless, the proposed method works in a non–strategic environment, where advertisers do not try to influence the outcome of the auction and always bid their true values. The strategic dimension of SSAs is partially taken into consideration in [16], where the advertisers are assumed to play a bidding strategy at the equilibrium of the GSP w.r.t. a set of estimated CTRs available to both the auctioneer and the advertisers. The authors introduce a learning algorithm which explores different rankings of the ads to improve the CTR estimates and, at the same time, to avoid that the advertisers have incentives to deviate from the aforementioned equilibrium strategy. In [17, 18], the authors formulate for the first time the problem of designing truthful learning mechanisms according to the notion of truthfulness in high probability in multi–slot SSAs. The single–slot online advertising is studied also in [19] where the notion of Bayesian Incentive Compatibility (BIC) is taken into consideration and an asymptotically BIC and *ex ante* efficient mechanism is introduced.

The most complete study of truthful bandit mechanisms so far is reported in [20] and [21]. These works provide a complete analysis on the constraints that truthfulness forces on the MAB algorithm with single–slot SSAs, proving that no *dominant–strategy* truthful bandit mechanism can achieve a regret (over social welfare or auctioneer's revenue) smaller than $\tilde{\Omega}(T^{\frac{2}{3}})$ and that the exploration and exploitation phases must be separate.[1,2] The lower bound over the regret holds also when the truthfulness is in expectation w.r.t. the click realizations. Finally, they also provide nearly–optimal

[1]The $\tilde{O}/\tilde{\Omega}/\tilde{\Theta}$ notation hides both constant and logarithmic factors, i.e., we say the regret is $\tilde{O}(T^{\frac{2}{3}})$ if there exist a and b such that the regret is $\leq aT^{\frac{2}{3}} \log^b T$.

[2]The need for having separated phases between exploration and exploitation to limit the strategic manipulation of the mechanism is underlined also in [16], where the authors study learning approaches for the GSP. Interestingly, experimental simulations show that having exploration phases in which no payment is applied can allow the auctioneer to have even a short–term gain [22].

algorithms matching the lower bound on the regret. In both [20] and [21], advertisers' utility is not subject to any form of time discount, in contrast with what happens in practice, where advertisers may favor early small gains over larger gains in the future. However, the mechanisms introduced in [20] and [21] are truthful even in presence of discount since the sharp separation of exploration and exploitation would still force advertisers with discounting to reveal their true valuation.[3] When the notion of truthfulness is relaxed, adopting truthfulness *in expectation* w.r.t. the mechanism randomness, it is possible to obtain in the case of single–slot SSAs a regret $\tilde{O}(T^{\frac{1}{2}})$ (over the social welfare) without separating the exploration and exploitation phases [23].

When multiple slots are present, a user model is needed to describe how the CTR of an ad changes as its position in the allocation changes. All the models available in the literature assume that the CTR is given by the product of two terms: the probability that an ad is clicked once observed by the user, and the probability that the user observes an ad given the complete allocation of ads over slots. The basic model (commonly referred to as *separability model*) prescribes that the probability of observing an ad depends only on its position [3]. Recently, more accurate models have been proposed and one of the most popular models is the *cascade model*. According to this model, the user scans the slots from top to bottom and the probability that she moves from a given slot to the next depends on the former slot itself and the identity of the ad displayed in it (this kind of user is commonly called *Markovian user*) [24, 25]. As a result, the overall probability of observing an ad depends on the slot in which it is displayed and on all the ads allocated above it. The validity of the cascade model has been evaluated and supported by a wide range of experimental investigations [26, 27]. The only results on learning mechanisms for SSAs with multiple slots are described in [28], where the authors characterize dominant–strategy truthful mechanisms and provide theoretical bounds over the social welfare regret for the separability model. However, these results are partial (e.g., they do not consider the common case in which the slot–dependent parameters are monotonically decreasing over slots), and they cannot be easily extended to the more challenging case of the cascade model (see Section 3.3).

[3] In our paper, we focus on the no–discount case and we use learning mechanisms that separate the phases of exploration and exploitation. As in [20] and [21], our learning mechanisms keep to be truthful even when discounting is present.

Original contributions. In the present paper, we build on the results available in the literature and we extend the partial results presented in [29] to a wider range of cases, providing also a number of contributions when the separability model and the cascade model are adopted. More precisely, our results can be summarized as follows.

- *Separability model with monotonically decreasing parameters/only position–dependent cascade model.* In this case, there are two groups of parameters, one related to the ads (called *quality*) and one to the slots (called *prominence*). We studied all the configurations of information incompleteness. When only qualities are unknown, we provide a non–randomized learning mechanism that is dominant–strategy truthful *a posteriori* w.r.t. the click realizations and with a regret of $\tilde{O}(T^{\frac{2}{3}})$ (while it is an open problem whether it is possible to obtain a better upper bound adopting truthfulness in expectation).[4] When only prominences are unknown, we provide a non–randomized learning mechanism that is dominant–strategy truthful in expectation w.r.t. the click realizations with a regret of 0 and a randomized learning mechanism that is dominant–strategy truthful in expectation w.r.t. the realizations of the random component of the mechanism with a regret of $O(1)$. We also show that any dominant–strategy truthful *a posteriori* w.r.t. all the sources of randomness learning mechanism would have a regret of $\Theta(T)$. When both groups of parameters are unknown, we provide a random learning mechanism that is dominant–strategy truthful in expectation only w.r.t. the realizations of the random component of the mechanism with a regret of $\tilde{O}(T^{\frac{2}{3}})$.

- *Cascade model*: in the non–factorized cascade model (i.e., when the observation probabilities can be arbitrary) we show that it is possible to obtain a regret of $\tilde{O}(T^{\frac{2}{3}})$ in dominant–strategy truthful in expectation w.r.t. all the sources of randomness learning mechanisms when only the qualities of the ads are unknown.[5] We show also that in the factorized cascade model (i.e., when the observation probabilities are the products

[4] This result has already been presented in [29] and is here reported for sake of completeness.

[5] A preliminary version of this result has already been presented in [29] and was here refined in its dependence from the number of slots K and ads N, changing from $\tilde{O}(T^{\frac{2}{3}}K^{\frac{2}{3}}N)$ to $\tilde{O}(T^{\frac{2}{3}}K^{\frac{4}{3}}N^{\frac{1}{3}})$, as postulated in the aforementioned paper.

of terms depending on either the slot or the ads as used in [24]), any non–randomized learning mechanism that is dominant–strategy truthful (even in expectation w.r.t. the click realizations) has a regret of $\Theta(T)$ even in the special case in which only the ad–dependent parameters are unknown (while it is an open problem whether it is possible to obtain a better upper bound adopting a randomized mechanism and truthfulness in expectation w.r.t. the realization of the random component of the mechanism).

- *Learning parameters*: for each setting described above we provide practical guidelines on how the learning parameters can be tuned to minimize the bound over the regret depending on the characteristics of the auction (e.g., number of slots and advertisers).

- *Numerical simulations*: we investigate the accuracy of all the theoretical regret bounds in predicting the dependency of the regret on the auction parameters by numerical simulations. We show that the theoretical asymptotic dependency matches the actual dependency we observed by simulation.

Paper organization. The paper is organized as follows. In Section 2, we briefly review the basics of mechanism design and MAB learning. In Section 3, we formalize SSAs, introduces the corresponding online learning mechanism design problem, and it provides a more formal overview of existing results and the new findings of this paper. In Sections 4 and 5, we report and discuss the main regret bounds in the case of position–dependent and position– and ad–dependent externalities. In Section 6, we report numerical simulations aiming at testing the accuracy of the theoretical bounds. Section 7 concludes the paper and proposes future directions of investigation. The detailed proofs of the theorems are reported in appendix.

2. Preliminaries

2.1. Economic Mechanisms

In this section we provide an overview on the definitions and results of mechanism design that are relevant to the paper. The objective of mechanism design [30] is to design *allocation* and *payment functions* satisfying some desirable properties when agents are *rational* and retain *private* information representing their preferences—also referred to as the *type* of the

agent. Without loss of generality, mechanism design focuses on specific mechanisms, called *direct*, in which the only action available to the agents is to report their (potentially non–truthful) type. On the basis of the agents' reports the mechanism determines the allocation of resources to agents and the agents' payments.

The main desirable property of a mechanism is *truthfulness*, often referred to as Incentive Compatibility (IC), which requires that reporting the true types constitutes an *equilibrium strategy profile* for the agents.[6] When a mechanism is not truthful, agents should try to optimize their (untruthful) strategies on the basis of some model about the opponents' behavior, but, in absence of common information, no normative model for rational agents exists. This leads the mechanism to be economically unstable, given that the agents continuously change their strategies. Different notions of truthfulness are available. The most common ones are Dominant Strategy Incentive Compatibility (DSIC)—i.e., reporting the true types is the best action an agent can play independently of the actions of the other agents—, *ex post incentive compatibility* (*ex post* IC)—i.e., reporting the true types is a Nash equilibrium—, and BIC—i.e., reporting the true types is a Bayes–Nash equilibrium. Interestingly, DSIC and *ex post* IC are equivalent notions of truthfulness in absence of interdependency among the types of the agents, while BIC is weaker than DSIC, since it only requires that every agent has a Bayesian prior over the types of the other agents and IC is defined in expectation w.r.t. the prior. When there are other sources of randomness in the mechanism design problem (not due to the distribution of probabilities over the types of the agents), e.g., random components of the mechanism or the realization of events, weaker solution concepts, said *in expectation*, are commonly adopted, e.g., DSIC in expectation or *ex post* IC in expectation. Instead, we use the term "*a posteriori*" when the truthfulness holds for every realization. In presence of multiple sources of randomness, a mechanism may be in expectation w.r.t. some sources and *a posteriori* w.r.t. other sources. When we use only DSIC *a posteriori* without specifying the source of randomness, we mean DSIC *a posteriori* w.r.t. all the sources of randomness. Moreover, mechanisms can exploit the realizations of the events adopting different payment functions for each different realization. These mechanisms are said Execution Contingent (EC) [31, 32].

[6] We use the same acronym also for 'Incentive Compatible' referred to a mechanism.

In addition to IC, other desirable properties include: Allocative Efficiency (AE)—i.e., the allocation maximizes the social welfare—, Individual Rationality (IR)—i.e., each agent is guaranteed to have no loss when reporting truthfully—, and Weak Budget Balance (WBB)—i.e., the mechanism is guaranteed to have no loss. In presence of sources of randomness, IR and WBB can be *in expectation* w.r.t. all the possible realizations, or *a posteriori* if they hold for every possible realization. As for IC, in presence of multiple sources of randomness, these properties may be in expectation w.r.t. some sources and *a posteriori* w.r.t. other sources. When we use only IR (or WBB) *a posteriori* without specifying the source of randomness, we mean IR (or WBB) *a posteriori* w.r.t. all the sources of randomness.

The economic literature provides an important characterization of the allocation functions that can be adopted in IC mechanisms when utilities are *quasi linear* [30]. Here, we survey the main results related to DSIC mechanisms where no sources of randomness are present. In unrestricted domains (i.e., the agents' types are defined over spaces with arbitrary structure) for the agents' preferences, only *weighted maximal–in–its–range* allocation functions can be adopted in DSIC mechanisms [33, 34]. More precisely, a weighted maximal–in–its–range allocation function chooses, among a subset of allocations that does not depend on the types reported by the agents (i.e., the range), the allocation maximizing the weighted social welfare, where each agent is associated with a positive (type–independent) weight. It trivially follows that, when the range is composed of all the possible allocations and all the agents have the same weights, only AE mechanisms can be DSIC. When weighted maximal–in–its–range allocation functions are adopted, only weighted Groves payments lead to DSIC mechanisms [30]. The most common DSIC mechanism is the VCG [30], in which the range is composed of all the allocations and all the weights are unitary. The idea of the VCG mechanism is that each agent pays the difference between the social welfare of the optimal outcome when she does not participate to the mechanism and the social welfare of the outcome obtained when she participates minute its contribution. (We leave a more formal and detailed description of the VCG mechanism to Appendix A.) Notice that the VCG mechanism satisfies also IR and WBB and, among all the Groves mechanisms, it is the one maximizing the revenue of the auctioneer. We refer to the weighted version of the VCG as Weighted Vickrey–Clarke–Groves (WVCG).

When the domain of the agents' preferences is restricted (i.e., the types are defined over spaces with specific structure, e.g., compact sets or discrete

values), weighted maximal–in–its–range property is not necessary for DSIC. The necessary condition is weakly monotonicity [30], which is also sufficient for convex domains. In specific restricted domains, weak monotonicity leads to simple and operational tools. For instance, when the preferences of the agents are single–parameter linear—i.e., the agents' value is given as the product between the agent's type and an allocation–dependent coefficient called *load* [35]—, monotonicity requires that the load is monotonically increasing in the type of the agent. In this case, any DSIC mechanism is based on the Myerson's payments defined in [35, 36].[7] Notice that the VCG mechanism is still the mechanism maximizing the auctioneer's revenue among all the DSIC mechanisms, including those that are not AE. The Myerson's payments include an integral that may be not easily computable. However, by adopting a random mechanism and accepting DSIC in expectation w.r.t. the realizations of the random component of the mechanism, such integral can be easily estimated by using samples [37]. Another drawback of the payments described in [35, 36] is that they require the off–line evaluation of the social welfare of the allocations for some agents' types different from the reported ones and this may be not possible in many practical situations. A way to overcome this issue is to adopt the result presented in [23], in which the authors propose an *implicit* way to calculate the payments. More precisely, given an allocation function in input, a random component is introduced such that with a small probability the reported types of the agents are modified to obtain the allocations that are needed to compute the payments in [35, 36]. The resulting allocation function is less efficient than the allocation function given in input, but the computation of the payments is possible and it is executed online.

2.2. Multi–Armed Bandit

The MAB [14] is a simple yet powerful framework formalizing the online decision–making problem under uncertainty. Historically, the MAB framework finds its motivation in optimal experimental design in clinical trials, where two treatments, say A and B, need to be tested. In an idealized version of the clinical trial, T patients are sequentially enrolled in the trial, so that whenever a treatment is tested on a patient, the outcome of the test

[7] See Appendix B for the definition of monotonicity in single–parameter linear environments and Myerson's payments.

is recorded and it is used to choose which treatment to provide to the next patient. The objective is to provide the best treatment to the largest number of patients. This raises the challenge of balancing the collection of information and the maximization of the performance of the trial, a problem usually referred to as the *exploration–exploitation* trade–off. On the one hand, it is important to gather information about the effectiveness of the two treatments by repeatedly providing them at different patients (*exploration*). On the other hand, as the estimate of effectiveness of the treatments becomes more accurate, the (estimated) best treatment should be selected more often (*exploitation*). This scenario matches with a large number of applications, such as online advertisements, adaptive routing, and cognitive radio.

In general, the MAB framework can be adopted whenever a set of N arms (e.g., treatments, ads) is available and the rewards (e.g., effectiveness of a treatment, CTR of an ad) associated with each of them are random realizations from unknown distributions. Although this problem can be solved by dynamic programming methods and notably by using the Gittins index solution [38], this requires a prior over the distribution of the reward of the arms and it is often computationally heavy (high–degree polynomial in T). More recently, a wide range of techniques have been developed to solve the bandit problem. In particular, these algorithms formalize the objective using the notion of *regret*, which corresponds to the difference in performance over T steps between an optimal selection strategy which knows in advance the performance of all the arms and an adaptive strategy which learns over time which arms to select. Although a complete review of the bandit algorithms is beyond the scope of this paper (see [39] for a review), we only discuss two results which are relevant to the rest of the paper. The *exploration–separated* algorithms solve the exploration–exploitation trade–off by introducing a strict separation between the exploration and the exploitation phases. While during the exploration phase all the arms are uniformly selected, in the exploitation phase only the best estimated arm is selected until the end of the experiment. The length τ of the exploration phase is critical to guarantee the success of the experiment and it is possible to show that, if properly tuned, the worst–case cumulative regret scales as $\tilde{O}(T^{\frac{2}{3}})$, matching the lower bound $\tilde{\Omega}(T^{\frac{2}{3}})$. Another class of algorithms relies on the construction of confidence intervals for the reward of each arm and it does not separate exploration and exploitation steps. In particular, the Upper–Confidence Bound (UCB) algorithm [40] gives an extra exploration bonus to

arms which have been selected only few times in the past and it achieves a worst–case cumulative regret of order $\tilde{O}(T^{\frac{1}{2}})$, matching the lower bound $\tilde{\Omega}(T^{\frac{1}{2}})$.

3. Problem statement

In this section we introduce all the notation used throughout the rest of the paper. In particular, we formalize the SSA model, we define the mechanism design problem, and we introduce the learning process.

Symbol	**Description**
N	Number of ads
$\mathcal{N} = \{1, \dots, N\}$	Set of the ads indexes
$a_i, i \in \mathcal{N}$	i-th ads
$q_i \in [0,1], i \in \mathcal{N}$	Quality for ad a_i
$\mathcal{V} = [0, V], V \in \mathbb{R}^+$	Set of the possible values/types for an ad
$v_i \in \mathcal{V}, i \in \mathcal{N}$	Value/types for ad a_i
$v_{\max} = \max_{i \in \mathcal{N}} v_i$	Maximum value
$\mathbf{v} = (v_1, \dots, v_N)$	Value profile
$\mathbf{v}_{-i} = (v_1, \dots, v_{i-1}, v_{i-1}, \dots, v_N)$	Value profile excluding the value for the i-th ad
$K, K < N$	Number of available slots
$\mathcal{K} = \{1, \dots, K\}$	Set of the available slots
$s_m, m \in \mathcal{K}$	m-th slot
$\mathcal{K}' = \mathcal{K} \bigcup \{K+1, \dots, N\}$	Extended set of the available slots
$\theta = \{\langle s_m, a_i \rangle : m \in \mathcal{K}', i \in \mathcal{N}\}$	Generic allocation
Θ	Set of all the possible allocations
$\pi : \mathcal{N} \times \Theta \to \mathcal{K}'$	Given an allocation θ, $\pi(i;\theta)$ returns the index of the slot in which a_i is allocated
$\alpha : \mathcal{K}' \times \Theta \to \mathcal{N}$	Given an allocation θ, $\alpha(m;\theta)$ returns the index of the ad allocated in slot s_m
$\gamma_{m,i}, m \in \mathcal{K}, i \in \mathcal{N}$	Probability that a user, observing ad a_i in slot s_m, observes the ad in the next slot s_{m+1}
$\Gamma_m(\theta), m \in \mathcal{K}, \theta \in \Theta$	Cumulative probability that a user observes the ad displayed at slot s_m in allocation θ

$\mathrm{SW}(\theta, \mathbf{v})$	Social welfare of allocation θ for ads with values profile $\mathbf{v}$
$\lambda_m \in [0,1], m \in \mathcal{K}$	Prominence associated with slot s_m
$c_i \in [0,1], i \in \mathcal{N}$	Continuation probability associated with ad a_i
$click_m^i(t) \in \{0,1\}$	No–click/click event for the ad a_i allocated in slot s_m at step t
$f : \mathcal{V}^N \to \Theta$	Allocation function
$p_i : \mathcal{V}^N \to \mathbb{R}$	Payment function for the i-th ad
$\hat{v}_i$	Reported value for ad a_i
$\hat{\mathbf{v}} = (\hat{v}_1, \ldots, \hat{v}_N)$	Reported value profile
$\hat{\mathbf{v}}_{-i} = (\hat{v}_1, \ldots, \hat{v}_{i-1}, \hat{v}_{i-1}, \ldots, \hat{v}_N)$	Reported value profile excluding the value for the i-th ad
θ^*	Allocation that maximizes the social welfare given the reported types
θ^*_{-i}	Allocation that maximizes the social welfare given the reported types when advertiser a_i is not present
$SW_{-i}(\theta, \mathbf{v})$	Cumulative expected value of the allocation θ minus the expected value of advertiser a_i
$R_T(\mathfrak{A})$	Expected revenue regret of algorithm $\mathfrak{A}$ over T steps
$R_T^{SW}(\mathfrak{A})$	Expected social welfare regret of algorithm $\mathfrak{A}$ over T steps

Table 1: Notation adopted throughout the paper.

3.1. SSA model

We resort to the standard model of SSAs [3]. The notation described in the sequel is summarized in Tab. 1. We denote by $\mathcal{N} = \{1, \ldots, N\}$ the set of ads indexes and by a_i with $i \in \mathcal{N}$ the i–th ad (we assume w.l.o.g. each advertiser has only one ad and therefore we can identify by a_i the i–th ad and the i–th advertiser indifferently). Each ad a_i is characterized by a *quality* q_i corresponding to the probability that a_i is clicked once observed by the user, and by the a *value* $v_i \in \mathcal{V} = [0, V]$ that a_i receives when clicked (a_i receives a value of zero if not clicked). We denote by $\mathbf{v}$ the profile $(v_1, \ldots, v_N)$

and by $\mathbf{v}_{-i}$ the profile obtained by removing v_i from $\mathbf{v}$. While the qualities $\{q_i\}_{i\in\mathcal{N}}$ may be known by the auctioneer with some level of accuracy, the values $\{v_i\}_{i\in\mathcal{N}}$ are private information of the advertisers. We denote by $\mathcal{K} = \{1, \ldots, K\}$ with $K < N$, the set of slot indexes and by s_m, with $m \in \mathcal{K}$, the m–th slot from top to bottom. For notational convenience, we also define the extended set of slots indexes $\mathcal{K}' = \mathcal{K} \cup \{K+1, \ldots, N\}$.[8]

We use the ordered pair $\langle s_m, a_i \rangle$ to indicate that ad a_i is allocated to slot s_m, while we denote by θ an *allocation*, defined as a collection of pairs $\langle s_m, a_i \rangle$, and by Θ the set of all the possible allocations. Although in an auction only K ads can be actually displayed, we define an allocation as $\theta = \{\langle s_m, a_i \rangle : m \in \mathcal{K}', i \in \mathcal{N}\}$ where both m and i occur exactly once and any ad assigned to a slot s_m with $m > K$ is not displayed. We define two maps $\pi : \mathcal{N} \times \Theta \to \mathcal{K}'$ and $\alpha : \mathcal{K}' \times \Theta \to \mathcal{N}$ such that $\pi(i;\theta)$ returns the index of the slot in which a_i is displayed in allocation θ and $\alpha(m;\theta)$ returns the index of the ad displayed in slot s_m in allocation θ. Given $\theta \in \Theta$, we have that $\pi(i;\theta) = m$ if and only if $\alpha(m;\theta) = i$.

With more than one slot, it is necessary to adopt a model of the user describing how the value of an advertiser varies over the slots. We assume that the user behaves according to the *cascade model* defined by [24, 25]. In the cascade model, the user's behavior is defined by a Markov chain whose possible states correspond to the slots, which are observed sequentially from the top to the bottom, and a transition matrix that defines, given the current slot, the probability that the user observes the ad a_i displayed in the next slot or stops observing any other ad. More precisely, the probability may depend on the index of the slot (i.e., $\pi(i;\theta)$), in this case the externalities are said *position–dependent*, and/or on the ad that precedes a_i in the current allocation θ (i.e., $a_{\alpha(\pi(i;\theta)-1;\theta)}$), in this case the externalities are said *ad–dependent*.

In the general case, the cascade model can be described by introducing a set of parameters $\gamma_{m,i}$ defined as the probability that a user, observing ad a_i in slot s_m, observes the ad in the next slot s_{m+1}. The probability that a user observes the ad displayed at slot s_m in allocation θ is denoted by $\Gamma_m(\theta)$ and

[8] Although $K < N$ is the most common case, the results could be smoothly extended to $K > N$.

it is defined as:

$$\Gamma_m(\theta) = \begin{cases} 1 & \text{if } m = 1 \\ \prod_{l=1}^{m-1} \gamma_{l,\alpha(l;\theta)} & \text{if } 2 \leq m \leq K \\ 0 & \text{otherwise} \end{cases} . \quad (1)$$

Given an allocation θ, the CTR of ad a_i is the probability to be clicked once allocated according to θ and it is computed as $\Gamma_{\pi(i;\theta)}(\theta)q_i$, corresponding to the joint probability that the user arrives at observing the slot in which the ad is displayed and then clicks on it. Similarly, the CTR of the ad displayed at slot s_m can be computed as $\Gamma_m(\theta)q_{\alpha(m;\theta)}$. We notice that, according to this model, the user may click multiple different ads at each impression. Given an allocation θ, the *expected value* (w.r.t. the click realizations) of advertiser a_i from θ is $\Gamma_{\pi(i;\theta)}(\theta)q_i v_i$, that is, the product of the CTR $\Gamma_{\pi(i;\theta)}(\theta)q_i$ by the value of the advertiser v_i. The advertisers' cumulative expected value from allocation θ, commonly referred to as *Social Welfare (SW)*, is:

$$\mathrm{SW}(\theta, \mathbf{v}) = \sum_{i=1}^{N} \Gamma_{\pi(i;\theta)}(\theta)q_i v_i.$$

In [24, 25], the authors factorize the probability $\gamma_{m,i}$ as the product of two independent terms: the *prominence* λ_m, which only depends on the slot s_m, and the *continuation probability* c_i, which only depends on the ad a_i.[9] In [24, 25], the authors factorize the probability $\gamma_{m,i}$ as the product of two independent terms: the *prominence* λ_m, which only depends on the slot s_m, and the *continuation probability* c_i, which only depends on the ad a_i.[10]

Finally, we denote by $click_m^i \in \{0, 1\}$ the no–click/click event for ad a_i allocated in slot s_m.

[9]The allocation problem when either all the prominence probabilities λ_ms or all the continuation probabilities c_is are equal to one can be solved in polynomial time, while, although no formal proof is known, the allocation problem with arbitrary λ_ms and c_is is commonly believed to be $\mathcal{NP}$–hard [24]. However, the allocation problem can be solved exactly in specific settings, and in many other cases, efficient approximation algorithms can be used [41]. In this paper, we ignore approximation schemes and we only focus on optimal allocation functions.

[10]The allocation problem when either all the prominence probabilities λ_ms or all the continuation probabilities c_is are equal to one can be solved in polynomial time, while, although no formal proof is known, the allocation problem with arbitrary λ_ms and c_is is commonly believed to be $\mathcal{NP}$–hard [24]. However, the allocation problem can be solved

3.2. Mechanism design problem

A direct–revelation economic mechanism for SSAs is formally defined as a tuple $(\mathcal{N}, \mathcal{V}, \Theta, f, \{p_i\}_{i\in\mathcal{N}})$ where $\mathcal{N}$ is the set of the agents' (i.e., the advertisers) indexes, $\mathcal{V}$ is the set of the types of the agents (where the type of ad a_i is the single–parameter valuation v_i), Θ is the set of the outcomes (i.e., the allocations), f is the allocation function defined as $f : \mathcal{V}^N \to \Theta$, and p_i is the payment function for advertiser a_i defined as $p_i : \mathcal{V}^N \to \mathbb{R}$. We denote by $\hat{v}_i$ the value reported by advertiser a_i to the mechanism, by $\hat{\mathbf{v}}$ the profile of reported values, and by $\hat{\mathbf{v}}_{-i}$ the profile obtained by removing $\hat{v}_i$ from $\hat{\mathbf{v}}$.

At the beginning of an auction, each advertiser a_i reports its value $\hat{v}_i$. The mechanism chooses the allocation on the basis of the values reported by the advertisers using $f(\hat{\mathbf{v}})$ and subsequently computes the payment of each advertiser a_i as $p_i(\hat{\mathbf{v}})$. The expected utility of advertiser a_i is defined as $\Gamma_{\pi(i;f(\hat{\mathbf{v}}))}(f(\hat{\mathbf{v}}))q_i v_i - p_i(\hat{\mathbf{v}})$, corresponding to the value expected by advertiser a_i minus the payment prescribed by the payment function. Notice that the utility is linear in the type of the agent. Since each advertiser is an expected utility maximizer, it will misreport its value (i.e., $\hat{v}_i \neq v_i$) whenever this may lead to increase its utility. Mechanism design aims at finding an allocation function f and a vector of payments $\{p_i\}_{i\in\mathcal{N}}$ such that some desirable properties—discussed in Section 2.1—are satisfied [30].

When the parameters q_i and $\gamma_{m,i}$ are known, the VCG mechanism satisfies DSIC in expectation w.r.t. the click realizations, IR in expectation w.r.t. the click realizations, WBB *a posteriori* w.r.t. the click realizations, and AE. DSIC and IR do not hold *a posteriori*. In the VCG mechanism, the allocation function, denoted by f^*, maximizes the SW given the reported types as

$$\theta^* = f^*(\hat{\mathbf{v}}) \in \arg\max_{\theta\in\Theta} \{\mathrm{SW}(\theta, \hat{\mathbf{v}})\} \tag{2}$$

and the payments are defined as

$$p_i^*(\hat{\mathbf{v}}) = \mathrm{SW}(\theta^*_{-i}, \hat{\mathbf{v}}_{-i}) - \mathrm{SW}_{-i}(\theta^*, \hat{\mathbf{v}}), \tag{3}$$

where:

- $\theta^*_{-i} = f^*(\hat{\mathbf{v}}_{-i})$, i.e., the optimal allocation when advertiser a_i is not present in the auction, and

exactly in specific settings, and in many other cases, efficient approximation algorithms can be used [41]. In this paper, we ignore approximation schemes and we only focus on optimal allocation functions.

- $\mathrm{SW}_{-i}(\theta^*, \hat{\mathbf{v}}) = \sum_{j \in \mathcal{N}, j \neq i} \Gamma_{\pi(j;\theta^*)}(\theta^*) q_j \hat{v}_j$, i.e., the cumulative expected value of the optimal allocation θ^* minus the expected value of advertiser a_i.

The payment of advertiser a_i is the difference between the SW that could be obtained from allocation θ^*_{-i}, computed removing ad a_i from the auction, and the SW of the efficient allocation θ^* without the contribution of advertiser a_i. In other words, this corresponds to the *cost* in terms of efficiency of the presence of a_i in the auction. The VCG mechanism can be easily extended to weighted case (the WVCG mechanism). The weighted SW is $\mathrm{SW}^w(\theta, \mathbf{v}) = \sum_{i=1}^{N} \Gamma_{\pi(i;\theta)}(\theta) q_i v_i w_i$ where w_i is the weight of advertiser i. In the WVCG, the allocation maximizing the weighted SW is chosen, while the payment is defined as $p_i^w(\hat{\mathbf{v}}) = \frac{1}{w_i}(\mathrm{SW}^w(\theta^*_{-i}, \hat{\mathbf{v}}_{-i}) - \mathrm{SW}^w_{-i}(\theta^*, \hat{\mathbf{v}}))$, where $\mathrm{SW}^w_{-i}(\theta^*, \hat{\mathbf{v}})) = \sum_{j \in \mathcal{N}, j \neq i} \Gamma_{\pi(j;\theta)}(\theta) q_j v_j w_j$.

The WVCG mechanism is DSIC in expectation w.r.t. the click realizations and IR in expectation w.r.t. the click realizations, but, WVCG being a generalization of the VCG, these properties do not hold *a posteriori*. This is because an advertiser may have a positive payment even when its ad has not been clicked. Nonetheless, the mechanism can be easily modified to satisfy DSIC w.r.t. the click realizations and IR *a posteriori* w.r.t. the click realizations by using *pay–per–click* payments $p_i^{*,c}$ as follows:

$$p_i^{*,c}(\hat{\mathbf{v}}, click^i_{\pi(i;\theta^*)}) = \frac{\mathrm{SW}(\theta^*_{-i}, \hat{\mathbf{v}}_{-i}) - \mathrm{SW}_{-i}(\theta^*, \hat{\mathbf{v}})}{\Gamma_{\pi(i;\theta^*)}(\theta^*) q_i} \times click^i_{\pi(i;\theta^*)}. \qquad (4)$$

The contingent formulation of the payments is such that $\mathbb{E}[p_i^{*,c}(\hat{\mathbf{v}}, click^i_{\pi(i;\theta^*)})] = p_i^*(\hat{\mathbf{v}})$, where the expectation is w.r.t. the click event, which is distributed as a Bernoulli random variable with parameter coinciding with the CTR of ad a_i in allocation θ^*, i.e., $\Gamma_{\pi(i;\theta^*)}(\theta^*) q_i$. Similar definitions hold for the WVCG mechanism.

3.3. Online learning mechanism design problem

In many practical problems, the parameters (i.e., q_i and $\gamma_{m,i}$) are not known in advance by the auctioneer and must be estimated at the same time as the auction is run. This leads to the definition of an iterative process where the auction is repeated over T steps using different estimates of the CTRs. This introduces a tradeoff between *exploring* different possible allocations, so as to collect information about the parameters, and *exploiting* the estimated parameters, so as to implement a truthful high–revenue auction (i.e., a VCG

mechanism). This problem could be easily cast as a MAB problem [14] and standard techniques could be used to solve it, e.g., [42]. Nonetheless, such an approach would completely overlook the strategic dimension of the problem: advertisers may choose their reported values at each step $t \in \{1 \dots, T\}$ to influence the outcome of the auction at t and/or in future steps in order to increase their cumulative utility over all the steps of the horizon T. Thus, in this context, truthfulness requires that reporting the truthful valuation maximizes the cumulative utility over the whole horizon T. The truthfulness can be in dominant strategies if advertisers know everything (including, e.g., the ads that will be clicked at each step t if displayed) or in expectation. As customary, we adopt four forms of truthfulness:

- DSIC *a posteriori* w.r.t. the click realizations and the realizations of the random component of the mechanism (if such a component is present),
- DSIC in expectation w.r.t. the click realizations and *a posteriori* w.r.t. the realizations of the random component of the mechanism (if such a component is present),
- DSIC in expectation w.r.t. the realizations of the random component of the mechanism and *a posteriori* w.r.t. the click realizations, and
- DSIC in expectation w.r.t. both randomization sources.

Here we face the challenging problem where the exploration–exploitation dilemma must be solved so as to maximize the revenue of the auction under the hard constraint of truthfulness. Let $\mathfrak{A}$ be a mechanism run over T steps. In particular, we only focus on mechanisms which are (at least) DSIC in expectation w.r.t. all sources of randomization, since for non-truthful mechanisms the dynamics of bids is unpredictable. At each step t, $\mathfrak{A}$ defines an allocation θ_t and prescribes an expected payment $p_{i,t}(\mathbf{v})$ for each ad a_i. The objective of $\mathfrak{A}$ is to minimize the loss of the auctioneer w.r.t. the revenue provided by the VCG mechanism computed on the actual parameters, and to preserve the properties of IR and WBB. More precisely, we measure the performance of $\mathfrak{A}$ as its cumulative expected regret over T steps:

$$R_T(\mathfrak{A}) = T \sum_{i=1}^{N} p_i^*(\mathbf{v}) - \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{N} p_{i,t}(\mathbf{v})\right], \tag{5}$$

where the expectation is taken w.r.t. the click realizations and the realization of the random component of the mechanism if present. Indeed, we remark that the regret is not defined on the basis of the pay–per–click payments asked on a specific sequence of clicks but on the expected payments $p_{i,t}(\mathbf{v})$. Furthermore, we notice that since the learning mechanism $\mathfrak{A}$ estimates the CTRs from the observed (random) clicks, the expected payments $p_{i,t}(\mathbf{v})$ are random as well. Finally, the payments are computed on the true valuations $\mathbf{v}$ constant over time since the mechanism $\mathfrak{A}$ is truthful by definition and thus the best option for all advertisers is to bid their true value at any step. The mechanism $\mathfrak{A}$ is a *no–regret* mechanism if its per–step regret $R_T(\mathfrak{A})/T$ decreases to 0 as T increases, i.e., $\lim_{T\to\infty} R_T(\mathfrak{A})/T = 0$. Another popular definition of performance [17, 21] is the SW regret, that measures the performance of $\mathfrak{A}$ as follows:

$$R_T^{SW}(\mathfrak{A}) = T \cdot \mathrm{SW}(\theta^*, \mathbf{v}) - \mathbb{E}\left[\sum_{t=1}^{T} \mathrm{SW}(\tilde{\theta}_t, \mathbf{v})\right], \tag{6}$$

where $\tilde{\theta}_t$ is the allocation prescribed by the learning mechanism at step t and the expectation, as before, is taken w.r.t. the click realizations and the realization of the random component of the mechanism if present. We notice that minimizing the SW regret R_T^{SW} does not coincide with minimizing R_T. In fact, once the quality estimates are accurate enough, such that θ_t is equal to θ^*, the SW regret drops to zero. On the other hand, since $p_{i,t}(\mathbf{v})$ is computed according to the estimated qualities, $R_T(\mathfrak{A})$ might still be positive even if $\theta_t = \theta^*$. In addition, we believe that in practical applications, providing a theoretical bound over the regret of the auctioneer's revenue is more important rather than a bound on the regret of the SW. Nevertheless, we show that the same approach we use to derive the bounds over the auctioneer's revenue can be employed to derive similar bounds over the SW. Finally, we refer to Appendix G for an alternative definition of the regret, related to the deviation between payments.

The study of the problem when $K = 1$ is well established in the literature. More precisely, the properties required to have a mechanism that is DSIC *a posteriori* w.r.t. the realizations of the random component of the mechanism are studied in [20] and it is shown that any learning algorithm must split the exploration and the exploitation in two separate phases in order to incentivize the advertisers to report their true values. This condition has a strong impact on the regret $R_T(\mathfrak{A})$ of the mechanism. In fact,

while in a standard bandit problem the distribution–free regret is of order $\Omega(T^{\frac{1}{2}})$, in single–slot auctions, DSIC *a posteriori* mechanisms have a regret $\Omega(T^{\frac{2}{3}})$. The same result holds for DSIC *a posteriori* w.r.t. the realizations of the random component of the mechanism and in expectation w.r.t. the click realizations. In [20], a truthful learning mechanism is designed with nearly optimal regret of order $\tilde{O}(T^{\frac{2}{3}})$. Similar structural properties for DSIC *a posteriori* w.r.t. the click realizations mechanisms are also studied in [21] and similar lower–bounds are derived for the SW regret. In [23] the authors show that, by introducing a random component in the allocation function and resorting to DSIC *a posteriori* w.r.t. the click realizations and in expectation w.r.t. the realizations of the random component of the mechanism, the separation of exploration and exploitation phases can be avoided. In this case, the upper bound over the regret of the SW is $\tilde{O}(T^{\frac{1}{2}})$, thus matching the best distribution–free bound in standard bandit problems. However, the payments of this mechanism suffer of potentially high variance, which may be an undesirable property in practice. Although we expect that this mechanism could also achieve a revenue regret of the order of $\tilde{O}(T^{\frac{1}{2}})$, no formal proof is known.

In this paper, we focus on the study of the problem when $K > 1$, which is still mostly unexplored. In this case, a crucial role is played by the CTR model. While with only one slot, the advertisers' CTRs coincide to their qualities q_i, with multiple slots the CTRs may also depend on the slots and the allocation of the other ads. The only results on learning mechanisms for SSAs with $K > 1$ are described in [28, 43], where the authors characterize DSIC *a posteriori* mechanisms and provide theoretical bounds over the SW regret. More precisely, the authors assume a simple CTR model in which the CTR itself depends on the ad a_i and the slot s_m. This model differs from the cascade model (see Section 2.1) where the CTR is a more complex function of the quality q_i of an ad and the cumulative probability of observation $\Gamma_m(\theta)$, which in general depends on both the slot s_m and the full allocation θ (i.e., the ads allocated before slot s_m). It can be easily shown that the model studied in [28] does not include and, at the same time, is not included by the cascade model. However, the two models match when the CTRs are separable in two terms, in which the first is the agents' quality and the second is a parameter in $[0, 1]$ monotonically decreasing over the slots (i.e., only–position–dependent cascade model). Furthermore, while the cascade model is supported by an empirical activity which confirms its validity as a model of

the user behavior [26, 27], the model considered in [28] has not been empirically studied. In [28], the authors show that when the CTRs are unrestricted (e.g., they are not strictly monotonically decreasing in the slots), the regret over the SW is $\Theta(T)$ and therefore at every step (of repetition of the auction) a non–zero regret is accumulated. In addition, the authors provide necessary and, in some situations, sufficient conditions to have DSIC *a posteriori* w.r.t. the click realizations. More precisely, the authors show that the allocation function of a mechanism that is DSIC *a posteriori* w.r.t. the click realizations must be monotonic *a posteriori* w.r.t. the click realizations. We recall that, given v_i and v_i' with $v_i' > v_i$ and called $\theta = f(v_i, \mathbf{v}_{-i})$ and $\theta' = f(v_i', \mathbf{v}_{-i})$, f is monotonic in expectation w.r.t. the click realizations if and only if the CTR of ad a_i in θ' is not strictly smaller than the CTR in θ. The definition of monotonicity *a posteriori* w.r.t. the click realizations is similar. Given v_i and v_i' with $v_i' > v_i$ and called $\theta = f(v_i, \mathbf{v}_{-i})$ and $\theta' = f(v_i', \mathbf{v}_{-i})$, f is monotonic *a posteriori* w.r.t. the click realizations if and only if the ad a_i is clicked in θ' whenever it would be clicked in θ. However, the authors do not present in [28] any bound over the regret (except for reporting an experimental evidence that the regret is $\Omega(T^{\frac{2}{3}})$ when the CTRs are separable). In [43], the authors preliminarily extend the analysis to the case of the cascade model, showing that, even with only ad–dependent externalities, any DSIC *a posteriori* w.r.t. the click realizations mechanism has a regret $\Theta(T)$.

We summarize in Tab. 2 the known results in the literature and, in bold font, the original results provided in this paper. We first consider the cascade model with only position–dependent externalities analyzing the case where only the parameters $\{q_i\}_{i \in \mathcal{N}}$ are unknown to the auctioneer. We show that it is possible to obtain a DSIC *a posteriori* learning mechanism with a regret $\tilde{\Theta}(T^{\frac{2}{3}})$ over the auctioneer's revenue.[11] Similarly, we show that in this setting, the regret over the SW is $\tilde{\Theta}(T^{\frac{2}{3}})$. In Section 4.2, we consider the opposite case where only the parameters $\{\Lambda_m\}_{m \in \mathcal{K}}$ are unknown. We focus on mechanisms that are DSIC in expectation w.r.t. the click realizations and *a posteriori* w.r.t. the random component of the mechanism in Section 4.2.1, and DSIC in expectation w.r.t. the realizations of the random component of the mechanism and *a posteriori* w.r.t. the click realizations in Section 4.2.2. In the first case we observe that we can obtain a mechanism

[11] A preliminary version of this result appears in [29].

slots	CTR model	unknown parameters	solution concept	regret over welfare (R_T^{SW})	regret over revenue (R_T)
1	–	q_i	DSIC	$\tilde{\Theta}(T^{2/3})$	$\tilde{\Theta}(T^{2/3})$
			DSICeC	$\tilde{\Theta}(T^{2/3})$	$\tilde{\Theta}(T^{2/3})$
			DSICeM	$\tilde{\Theta}(T^{1/2})$	$\tilde{O}(T^{2/3})$
> 1	(unconstrained) $CTR_{i,m}$	$CTR_{i,m}$	DISC	$\Theta(T)$	unknown
	(unfactorized) cascade	q_i	DISC	$\Theta(T)$	$\Theta(T)$
			DSICeM	$\Theta(T)$	$\Theta(T)$
			DSICeC	$\boldsymbol{\tilde{\Theta}(\mathbf{T^{2/3}})}$	$\boldsymbol{\tilde{\Theta}(\mathbf{T^{2/3}})}$
		$\gamma_{i,s}$	DISCeM	$\boldsymbol{\Theta(\mathbf{T})}$	$\boldsymbol{\Theta(\mathbf{T})}$
	position–dep. cascade / separable $CTR_{i,m}$	q_i	DSIC	$\boldsymbol{\tilde{\Theta}(\mathbf{T^{2/3}})}$	$\boldsymbol{\tilde{\Theta}(\mathbf{T^{2/3}})}$
			DSICeC	$\boldsymbol{\tilde{\Theta}(\mathbf{T^{2/3}})}$	$\boldsymbol{\tilde{\Theta}(\mathbf{T^{2/3}})}$
			DSICeM	$\boldsymbol{\tilde{O}(\mathbf{T^{2/3}})}$	$\boldsymbol{\tilde{O}(\mathbf{T^{2/3}})}$
		λ_m	DSIC	$\boldsymbol{\Theta(\mathbf{T})}$	$\boldsymbol{\Theta(\mathbf{T})}$
			DSICeC	$\mathbf{0}$	$\mathbf{0}$
			DSICeM	$\boldsymbol{O(1)}$	$\boldsymbol{O(1)}$
		q_i, λ_m	DSIC	$\boldsymbol{\Theta(\mathbf{T})}$	$\boldsymbol{\Theta(\mathbf{T})}$
			DSICeM	$\boldsymbol{\tilde{O}(\mathbf{T^{2/3}})}$	$\boldsymbol{\tilde{O}(\mathbf{T^{2/3}})}$
	ad–dependent cascade	q_i	DSIC	$\Theta(T)$	$\Theta(T)$
			DSICeC	$\boldsymbol{\tilde{\Theta}(\mathbf{T^{2/3}})}$	$\boldsymbol{\tilde{\Theta}(\mathbf{T^{2/3}})}$
		c_i	DSICeC	$\boldsymbol{\Theta(\mathbf{T})}$	$\boldsymbol{\Theta(\mathbf{T})}$
		q_i, c_i	DSICeC	$\boldsymbol{\Theta(\mathbf{T})}$	$\boldsymbol{\Theta(\mathbf{T})}$

Table 2: Results available on the regret of learning mechanisms in SSA, with in bold the original results derived in this paper. 'DSIC' stands for 'DSIC *a posteriori*'; 'DSICeC' stands for 'DSIC in expectation w.r.t. the click realizations and *a posteriori* w.r.t. realizations of the random component of the mechanism'; 'DSICeM' stands for 'DSIC in expectation w.r.t. realizations of the random component of the mechanism and *a posteriori* w.r.t. the click realizations'.

with a regret (both over the auctioneer's revenue and over the SW) of 0, but the obtained mechanism is WBB only in expectation w.r.t. the click realizations. In the second scenario, both the regrets are bounded by a constant and the mechanism is IR *a posteriori* and WBB in expectation w.r.t. the random component of the mechanism. In Section 4.2.3, we derive a negative result on the possibility of having no–regret DSIC *a posteriori* w.r.t. both sources of randomness mechanisms. Obviously, this negative result extends to all the generalizations of the cascade model with only position–dependent externalities. We conclude the analysis of the position–dependent model studying, in Section 4.3, the case where both $\{\Lambda_m\}_{m\in\mathcal{K}}$ and $\{q_i\}_{i\in\mathcal{N}}$ are unknown by the auctioneer, showing that it is possible to obtain DSIC in expectation w.r.t.

the random component of the mechanism and *a posteriori* w.r.t. the click realizations mechanisms with bounds of $\tilde{O}(T^{\frac{2}{3}})$ for both kinds of regret.

In Section 5 we study the cascade model with both position– and ad–dependent externalities. We provide a DSIC in expectation w.r.t. the click realizations learning algorithm minimizing the regret over the auctioneer's revenue when only the parameters $\{q_i\}_{i\in\mathcal{N}}$ are unknown.[12] Then we provide a result over the SW regret, where the bound is still $\tilde{\Theta}(T^{\frac{2}{3}})$. Finally, in Section 5.2, we consider other situations of lack of information obtaining negative results. More precisely, there is not any no–regret DSIC mechanism in expectation w.r.t. the click realizations and *a posteriori* w.r.t. realizations of the random component of the mechanism when parameters $\{c_i\}_{i\in\mathcal{N}}$ are unknown. This result applies to both kinds of regret and it extends to the more general cascade model.

4. Learning with Position–Dependent Externalities

In this section we study the multi–slot auctions with only position–dependent cascade model. The CTRs depend only on the quality of the ads and on the position of the slots in which the ads are allocated. Formally, the parameters $\gamma_{m,i}$ coincide with the prominence parameter, i.e., $\gamma_{m,i} = \lambda_m$ for every $m \in \mathcal{K}$ and $i \in \mathcal{N}$. As a result, the cumulative probability of observation, defined in Eq. 1, reduces to

$$\Lambda_m = \Gamma_m(\theta) = \begin{cases} 1 & \text{if } m = 1 \\ \prod_{l=1}^{m-1} \lambda_l & \text{if } 2 \leq m \leq K \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

where we use Λ_m instead of $\Gamma_m(\theta)$ for consistency with most of the literature on position–dependent externalities and to stress the difference w.r.t. to the general case described in Section 3.1.

When all the parameters are known by the auctioneer, the efficient allocation θ^* greatly simplifies. In fact, the expected value of an ad a_i for an allocation θ in this case reduces to $\Lambda_{\pi(i;\theta)} q_i v_i$ and, since the cumulative probabilities of observation are non–increasing over slots, the efficient allocation

[12] A preliminary version of this result appears in [29]. In the current paper we provide a more accurate bound filling the gap between the dependence over N and K predicted by the theoretical bound and the results in the numerical simulation.

simply needs to allocate the slots in decreasing order of their reported values in expectation w.r.t. the qualities, i.e., $q_i\hat{v}_i$. More precisely, for any $k \in \mathcal{K}'$, let $\max_{i\in\mathcal{N}}(q_i\hat{v}_i; k)$ be the operator returning the k–th largest value in the set $\{q_1\hat{v}_1, \ldots, q_N\hat{v}_N\}$, then θ^* is such that, for every $m \in \mathcal{K}'$, the ad displayed at slot s_m is

$$\alpha(m; \theta^*) \in \arg\max_{i\in\mathcal{N}}(q_i\hat{v}_i; m). \tag{8}$$

This condition also simplifies the definition of the efficient allocation θ^*_{-i} when a_i is removed from $\mathcal{N}$. In fact, for any $i, j \in \mathcal{N}$, if $\pi(j; \theta^*) < \pi(i; \theta^*)$ (i.e., ad a_j is displayed before a_i) then $\pi(j; \theta^*_{-i}) = \pi(j; \theta^*)$, while if $\pi(j; \theta^*) > \pi(i; \theta^*)$ then $\pi(j; \theta^*_{-i}) = \pi(j; \theta^*)-1$ (i.e., ad a_j is moved one slot upward), and w.l.o.g. we assume $\pi(i; \theta^*_{-i}) = N$. In case of position–dependent externalities, the VCG payments p^*_i defined in Eq. 3 take the simplified formulation:

$$p^*_i(\hat{\mathbf{v}}) = \begin{cases} \sum_{l=\pi(i;\theta^*)+1}^{K+1} \left[(\Lambda_{l-1} - \Lambda_l) \max_{j\in\mathcal{N}}(q_j\hat{v}_j; l) \right] & \text{if } \pi(i; \theta^*) \leq K \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

which can be written as a per–slot payment as:

$$p^*_{\alpha(m;\theta^*)}(\hat{\mathbf{v}}) = \begin{cases} \sum_{l=m+1}^{K+1} \left[(\Lambda_{l-1} - \Lambda_l) \max_{i\in\mathcal{N}}(q_i\hat{v}_i; l) \right] & \text{if } m \leq K \\ 0 & \text{otherwise} \end{cases}. \tag{10}$$

In the following sections we study the problem of designing incentive compatible mechanisms under different conditions of lack of information over the parameters $\{q_i\}_{i\in\mathcal{N}}$ and $\{\Lambda_m\}_{m\in\mathcal{K}}$. In particular, in Section 4.1, we assume that the actual values of $\{q_i\}_{i\in\mathcal{N}}$ are unknown by the auctioneer, while those of $\{\Lambda_m\}_{m\in\mathcal{K}}$ are known. In Section 4.2, we assume that the actual values of $\{\Lambda_m\}_{m\in\mathcal{K}}$ are unknown by the auctioneer, while those of $\{q_i\}_{i\in\mathcal{N}}$ are known. Finally, in Section 4.3, we assume that both $\{q_i\}_{i\in\mathcal{N}}$ and $\{\Lambda_m\}_{m\in\mathcal{K}}$ are unknown.

4.1. Unknown qualities $\{q_i\}_{i\in\mathcal{N}}$

In this section we assume that the qualities of the ads $\{q_i\}_{i\in\mathcal{N}}$ are unknown, while $\{\Lambda_m\}_{m\in\mathcal{K}}$ are known. We initially focus on DSIC *a posteriori*

Input: Length of exploration phase τ, confidence δ, position–dependent parameters $\{\Lambda_m\}_{m\in\mathcal{K}}$

Exploration phase
for $t = 1, \ldots, \tau$ **do**
  Allocate ads according to (13)
  Ask for no payment
  Observe the clicks $\{click^i_{\pi(i;\theta_t)}(t)\}_{i=1}^N$
Compute the estimated quality $\tilde{q}_i = \frac{1}{|B_i|}\sum_{t\in B_i} \frac{click^i_{\pi(i;\theta_t)}(t)}{\Lambda_{\pi(i;\theta_t)}}$
Compute $\tilde{q}_i^+ = \tilde{q}_i + \eta$ where η is given by (14)

Exploitation phase
for $t = \tau+1, \ldots, T$ **do**
  Allocate ads according to $\tilde{f}$ defined in (16)
  For each ad a_i, ask for payment $\tilde{p}_i^c$ defined in (18)

Figure 1: Pseudo–code for the A–VCG1 mechanism.

mechanisms and subsequently we discuss about DSIC in expectation mechanisms.

As in [20, 21], we formalize the problem as a MAB problem and we study the properties of a learning mechanism where the exploration and exploitation phases are separated, such that during the exploration phase, we estimate the values of $\{q_i\}_{i\in\mathcal{N}}$ and during the exploitation phase we use the estimated qualities $\{\tilde{q}_i\}_{i\in\mathcal{N}}$ to implement a DSIC *a posteriori* mechanism. The pseudo code of the algorithm A–VCG1 (Adaptive VCG 1) is given in Fig. 1. The details of the algorithm follow.

Exploration phase. During an exploration phase of length τ, the algorithm receives as input the parameters $\{\Lambda_m\}_{m\in\mathcal{K}}$ and collects data to estimate the quality of each ad. Unlike the single–slot case, where we collect only one sample of no–click/click event per step, here we can exploit the fact that each ad a_i has a non–zero CTR whenever it is allocated to a slot s_m with $m \leq K$. As a result, at each step of the exploration phase, we collect K samples (no–click/click events), one from each slot. Let θ_t (for $1 \leq t \leq \tau$) be a sequence of allocations independent from the advertisers' bids. Let $B_i = \{t : \pi(i;\theta_t) \leq K, \leq t \leq \tau\}$ be the set of all the steps when a_i is

allocated to a valid slot, so that $|B_i|$ corresponds to the total number of (no–click/click) samples available for ad a_i. We denote by $click^i_{\pi(i;\theta_t)}(t) \in \{0,1\}$ the no–click/click event at step t for ad a_i when displayed in slot $s_{\pi(i;\theta_t)}$. Depending on the slot in which the click event happens, the ad a_i has different CTR, thus we need to weigh each click sample by the probability of observation Λ_m related to the slot in which the ad is allocated. As a result, the estimated quality $\tilde{q}_i$ is computed as

$$\tilde{q}_i = \frac{1}{|B_i|} \sum_{t \in B_i} \frac{click^i_{\pi(i;\theta_t)}(t)}{\Lambda_{\pi(i;\theta_t)}}. \tag{11}$$

Since $\tilde{q}_i$ is an unbiased estimate of q_i (i.e., $\mathbb{E}_{click}[\tilde{q}_i] = q_i$, where $\mathbb{E}_{click}$ is the expectation w.r.t. the click realizations), we can resort to the Hoeffding's inequality [44] and a bound over the error of the estimated quality $\tilde{q}_i$ for each ad a_i.

Proposition 1. *For any ad a_i, $i \in \mathcal{N}$*

$$|q_i - \tilde{q}_i| \leq \sqrt{\left(\sum_{t \in B_i} \frac{1}{\Lambda^2_{\pi(i;\theta_t)}} \right) \frac{1}{2|B_i|^2} \log \frac{2N}{\delta}}, \tag{12}$$

with probability $1 - \delta$ (w.r.t. the click realizations).

During the exploration phase, at each step $t \in \{1, \dots, \tau\}$, we adopt the following sequence of allocations

$$\theta_t = \{\langle s_1, a_{(t \bmod N)+1} \rangle, \dots, \langle s_N, a_{(t+N-1 \bmod N)+1} \rangle\}, \tag{13}$$

thus obtaining $|B_i| = \lfloor K\tau/N \rfloor$ for all the ads a_i. Given that $\lfloor K\tau/N \rfloor \geq \frac{\tau K}{2N}$, Eq. 12 becomes

$$|q_i - \tilde{q}_i| \leq \sqrt{\left(\sum_{m=1}^{K} \frac{1}{\Lambda^2_m} \right) \frac{N}{K^2 \tau} \log \frac{2N}{\delta}} =: \eta, \tag{14}$$

where η represents the accuracy of the estimator.[13] The previous inequality is non-trivial only for a long enough exploration phase. In particular, to

[13]Notice that, from now on, we realistically assume that all the ads have at least two samples to initialize their estimates $\tilde{q}^+_i$. This hypothesis allows us to remove the floor notation in the bounds and, in the case of A–VCG1, it leads to an exploration time $\tau \geq 2N/K$.

have a meaningful bound, i.e., $|q_i - \tilde{q}_i| < 1$, the length of the exploration phase has to be $\tau > \left(\sum_{m=1}^{K} \frac{1}{\Lambda_m^2}\right) \frac{N}{K^2} \log \frac{2N}{\delta}$. During this phase, in order to guarantee DSIC *a posteriori*, the advertisers cannot be charged with any payment, i.e., all the payments in steps $1 \leq t \leq \tau$ are set to 0. In fact, as shown in [21], any bid–dependent payment could be easily manipulated by bidders, thus obtaining a non–truthful mechanism, whereas bid–independent payments could lead to a non–IR mechanism to which bidders may prefer not to participate.

Exploitation phase. Once the exploration phase is terminated, an upper–confidence bound over each quality q_i is computed as

$$\tilde{q}_i^+ = \tilde{q}_i + \eta, \tag{15}$$

and the exploitation phase is started and run for the remaining $T - \tau$ steps. We define the upper-confidence bound on the SW as

$$\widetilde{\mathrm{SW}}(\theta, \hat{\mathbf{v}}) = \sum_{i=1}^{N} \Lambda_{\pi(i;\theta)} \tilde{q}_i^+ \hat{v}_i,$$

and we define $\tilde{f}$ as the allocation function that displays ads in decreasing order of $\tilde{q}_i^+ \hat{v}_i$. Thus $\tilde{f}$ returns the efficient allocation $\tilde{\theta}$ on the basis of the estimated qualities as:

$$\tilde{\theta} = \tilde{f}(\hat{\mathbf{v}}) \in \arg\max_{\theta \in \Theta} \ \{\widetilde{\mathrm{SW}}(\theta, \hat{\mathbf{v}})\}. \tag{16}$$

The allocation function $\tilde{f}$ is then run for all the steps of the exploitation phase. Notice that $\tilde{f}$ is an affine maximizer, since

$$\begin{aligned} \tilde{f}(\hat{\mathbf{v}}) \in \arg\max_{\theta \in \Theta} \sum_{i=1}^{N} \Lambda_{\pi(i;\theta)} \tilde{q}_i^+ \hat{v}_i &= \arg\max_{\theta \in \Theta} \sum_{i=1}^{N} \frac{\tilde{q}_i^+}{q_i} \Lambda_{\pi(i;\theta)} q_i \hat{v}_i \\ &= \arg\max_{\theta \in \Theta} \sum_{i=1}^{N} w_i \Lambda_{\pi(i;\theta)} q_i \hat{v}_i, \end{aligned}$$

where each weight $w_i = \tilde{q}_i^+ / q_i$ is independent of the advertisers' types v_i. Hence, we can apply the WVCG payments (here denoted by $\tilde{p}$ because of

the estimated parameters) satisfying the DSIC *a posteriori* property. In particular, for any a_i, we define the payment

$$\begin{aligned}\tilde{p}_i(\hat{\mathbf{v}}) &= \begin{cases} \frac{1}{w_i}\sum_{l=\pi(i;\tilde{\theta})+1}^{K+1}(\Lambda_{l-1}-\Lambda_l)\max_{j\in\mathcal{N}}(\tilde{q}_j^+\hat{v}_j;l) & \text{if } \pi(i;\tilde{\theta})\leq K \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{q_i}{\tilde{q}_i^+}\sum_{l=\pi(i;\tilde{\theta})+1}^{K+1}(\Lambda_{l-1}-\Lambda_l)\max_{j\in\mathcal{N}}(\tilde{q}_j^+\hat{v}_j;l) & \text{if } \pi(i;\tilde{\theta})\leq K \\ 0 & \text{otherwise} \end{cases}. \end{aligned} \tag{17}$$

Although these payments cannot be computed by the auctioneer, since the actual $\{q_i\}_{i\in\mathcal{N}}$ are unknown, we can resort to the *pay–per–click* payments

$$\begin{aligned}\tilde{p}_i^c(\hat{\mathbf{v}},click^i_{\pi(i;\tilde{\theta})}) &= \frac{\tilde{p}_i(\hat{\mathbf{v}})}{\Lambda_{\pi(i;\tilde{\theta})}q_i} \\ &= \frac{1}{\Lambda_{\pi(i;\tilde{\theta})}\tilde{q}_i^+}\Bigg(\sum_{l=\pi(i;\tilde{\theta})+1}^{K+1}(\Lambda_{l-1}-\Lambda_l)\max_{j\in\mathcal{N}}(\tilde{q}_j^+\hat{v}_j;l)\Bigg)click^i_{\pi(i;\tilde{\theta})}, \end{aligned} \tag{18}$$

which in expectation coincide with the payments $\tilde{p}_i(\hat{\mathbf{v}}) = \mathbb{E}[\tilde{p}_i^c(\hat{\mathbf{v}}, click^i_{\pi(i;\tilde{\theta})})]$ and can be computed at run time. Unlike the payments $\tilde{p}_i(\hat{\mathbf{v}})$, these payments can be computed simply relying on the estimates $\tilde{q}_i^+$ and on the knowledge of the probabilities of observation Λ_m. The following properties hold for this mechanism.

Theorem 1. *The A–VCG1 is:*

- *DSIC* a posteriori,
- *IR* a posteriori,
- *WBB* a posteriori.

Proof. The allocation function is monotonic *a posteriori* w.r.t. the click realizations since, by the nature of the externality model, each click realization plan prescribing that an ad is clicked in a given slot prescribes also that the same ad would be clicked in all the slots above. Thus, DSIC *a posteriori* trivially follows from the fact that the mechanism is a WVCG mechanism and that the payments are pay–per–click. □

We now move to the analysis of the performance of A–VCG1 in terms of the regret the mechanism accumulates through T steps.

Theorem 2. *Let us consider a sequential auction with N advertisers, K slots, and T steps with position–dependent cascade model with parameters $\{\Lambda_m\}_{m=1}^K$ and accuracy η as defined in Eq. 14. For any parameter $\tau \in \{1,\dots,T\}$ and $\delta \in (0,1)$, the A–VCG1 achieves auctioneer's revenue expected regret:*

$$R_T \leq v_{\max}\left(\sum_{m=1}^{K}\Lambda_m\right)\Big(2(T-\tau)\eta + \tau + \delta T\Big). \tag{19}$$

By setting the parameters to

- $\delta = K^{-\frac{1}{3}}T^{-\frac{1}{3}}N^{\frac{1}{3}}$
- $\tau = K^{-\frac{1}{3}}T^{\frac{2}{3}}N^{\frac{1}{3}}\Lambda_{\min}^{-\frac{2}{3}}\left[\log\left(K^{\frac{1}{3}}T^{\frac{1}{3}}N^{\frac{2}{3}}\right)\right]^{\frac{1}{3}}$,

where $\Lambda_{\min} = \min_{m\in\mathcal{K}}\Lambda_m > 0$, then the regret is

$$R_T \leq 4v_{\max}\Lambda_{\min}^{-\frac{2}{3}}K^{\frac{2}{3}}T^{\frac{2}{3}}N^{\frac{1}{3}}\left[\log\left(K^{\frac{1}{3}}T^{\frac{1}{3}}N^{\frac{2}{3}}\right)\right]^{\frac{1}{3}}. \tag{20}$$

We initially introduce some remarks about the above results, and subsequently discuss the proof of the theorem, which can be found in the appendix.

Remark 1 (The bound). Up to numerical constants and logarithmic factors, the bound in Eq. 20 is $R_T = \tilde{O}(K^{\frac{2}{3}}T^{\frac{2}{3}}N^{\frac{1}{3}})$. We first notice that A–VCG1 is a no–regret algorithm since its per–step regret (R_T/T) decreases to 0 as $T^{-\frac{1}{3}}$, thus implying that it asymptotically achieves the same performance as the VCG. Furthermore, we notice that for $K = 1$ the bound reduces (up to constants) to the single–slot case analyzed in [20]. Unlike the standard bound for MAB algorithms, the regret scales as $\tilde{O}(T^{\frac{2}{3}})$ instead of $\tilde{O}(T^{\frac{1}{2}})$. As pointed out in [20] and [21] this is the unavoidable price the bandit algorithm has to pay to be DSIC *a posteriori* w.r.t. the realizations of the random component of a mechanism. Finally, the dependence of the regret on N is sub–linear ($N^{\frac{1}{3}}$) and therefore increasing the number of advertisers does not significantly worsen the regret. The dependency on the number of slots K is similar: according to the bound in Eq. 20 the regret has a sublinear dependency $\tilde{O}(K^{\frac{2}{3}})$, meaning that whenever one slot is added to the auction,

the performance of the algorithm does not significantly worsen. By analyzing the difference between the payments of the VCG and A–VCG1, we notice that during the exploration phase the regret is $O(\tau K)$ (e.g., if all the ads allocated into the K slots are clicked at each explorative step), while during the exploitation phase the error in estimating the qualities sum over all the K slots, thus suggesting a linear dependency on K for this phase as well. Nonetheless, as K increases, the number of samples available per–ad increases as $\tau K/N$, thus improving the accuracy of the quality estimates by $\tilde{O}(K^{-\frac{1}{2}})$ (see Proposition 1). As a result, as K increases, the exploration phase can be shortened (the optimal τ actually decreases as $K^{-\frac{1}{3}}$), thus reducing the regret during the exploration, and still have accurate enough estimates to control the regret of the exploitation phase.

Remark 2 (Distribution–free bound). The bound derived in Theorem 2 is a *distribution–free* (or worst–case) bound, since it holds for any set of advertisers (i.e., for any $\{q_i\}_{i\in\mathcal{N}}$ and $\{v_i\}_{i\in\mathcal{N}}$). This generality comes at the price that, as illustrated in other remarks and in the numerical simulations (see Section 6), the bound could be inaccurate for some specific sets of advertisers. On the other hand, distribution–dependent bounds (see e.g., the bounds of UCB [42]), where q and v appear explicitly, would be more accurate in predicting the behavior of the algorithm. Nonetheless, they could not be used to optimize the parameters δ and τ, since they would then depend on unknown quantities.

Remark 3 (Parameters). The choice of parameters τ and δ reported in Theorem 2 is obtained by a rough minimization of the upper–bound in Eq. 19. Each parameter can be computed by knowing the characteristics of the auction (number of steps T, number of slots K, number of ads N, and Λ_m). Moreover, since the values are obtained optimizing an upper–bound of the regret and not directly the true cumulative regret, these values can provide a good guess for the parametrization, but they might not be optimal. In practice, we expect that the regret could be optimized by searching the space of the parameters around the values suggested in Theorem 2.

Remark 4 (DSIC in expectation). In this paper, we do not solve two interesting problems when DSIC in expectation w.r.t. the realizations of the random component of the mechanism is adopted: *(i)* whether it is possible or not to avoid the separation of the exploration and exploitation phases and *(ii)* whether it is possible to obtain a regret of $\tilde{O}(T^{\frac{1}{2}})$ as in the case of $K = 1$ [23]. Any attempt we tried to extend the result presented in [23]

to the multi–slot case led to a non–IC mechanism. We briefly provide some examples of adaptation to our framework of the two MAB presented in [23]. None of these attempts provided a monotonic allocation function. We tried to extend the UCB1 in different ways, e.g., by introducing $N \cdot K$ estimators, one for each ad for each slot, or maintaining N estimators weighing in different ways clicks obtained in different slots. The second MAB algorithm, called NewCB, is based on the definition of a set of active ads, i.e., the ones that can be displayed in the unique slot. We considered some extensions for the multi–slot setting (e.g. a single set for all the slots and multiple sets, one for each slot) without identifying monotonic allocation algorithms.

Comments to the proof. The proof uses relatively standard arguments to bound the regret of the exploitation phase. As discussed in Remark 2, the bound is distribution–free and some steps in the proof are conservative upper–bounds on quantities that might be smaller for specific auctions. For instance, the inverse dependency on the smallest cumulative discount factor $\Lambda_{\min}$ in the final bound could be a quite inaccurate upper–bound on the quantity $\sum_{m=1}^{K} 1/\Lambda_m^2$. In fact, the parameter τ itself could be optimized as a direct function of $\sum_{m=1}^{K} 1/\Lambda_m^2$, thus obtaining a more accurate tuning of the length of the exploration phase and a slightly tighter bound (in terms of constant terms). Furthermore, a crucial step in the proof is the inequality $\max_{i \in \mathcal{N}}(\tilde{q}_i^+ v_i; h) / \max_{i \in \mathcal{N}}(\tilde{q}_i^+ v_i; m) \leq 1$, which is likely to become less accurate as the difference between h and m increases (see Eq. C.4 in the proof). For instance, if the qualities q_i are drawn from a uniform distribution in $(0, 1)$, as the number of slots increases this quantity reduces as well (on average), thus making the upper–bound by 1 less and less accurate. The accuracy of the proof and the corresponding bound are further studied in the simulations in Section 6.

Besides a bound on the revenue regret, in a similar way we can bound the SW, as in [23]. In particular, we obtain that A–VCG1 is a no–regret algorithm and $R_T^{SW} = \tilde{O}(T^{\frac{2}{3}})$.

Theorem 3. *Let us consider a sequential auction with N advertisers, K slots, and T steps with position–dependent cascade model with parameters $\{\Lambda_m\}_{m=1}^{K}$ and accuracy η as defined in Eq. 14. For any parameter $\tau \in \{1, \dots, T\}$ and $\delta \in (0, 1)$, the A–VCG1 achieves a SW regret:*

$$R_T^{SW} \leq v_{\max} K \left(2 \left(T - \tau\right) \eta + \tau + \delta T\right). \tag{21}$$

By setting the parameters to

- $\delta = \left(\frac{1}{\Lambda_{\min}}\right)^{\frac{2}{3}} K^{-\frac{1}{3}} T^{-\frac{1}{3}} N^{\frac{1}{3}}$
- $\tau = \left(\frac{1}{\Lambda_{\min}}\right)^{\frac{2}{3}} K^{-\frac{1}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log 2\Lambda_{\min}^{\frac{2}{3}} K^{\frac{1}{3}} T^{\frac{1}{3}} N^{\frac{2}{3}}\right)^{\frac{1}{3}}$,

where $\Lambda_{\min} = \min_{m \in \mathcal{K}} \Lambda_m,\ \Lambda_{\min} > 0$, *then the regret is*

$$R_T^{SW} \leq 4 v_{\max} \left(\frac{1}{\Lambda_{\min}}\right)^{\frac{2}{3}} K^{\frac{2}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log 2\Lambda_{\min}^{\frac{2}{3}} K^{\frac{1}{3}} T^{\frac{1}{3}} N^{\frac{2}{3}}\right)^{\frac{1}{3}}. \tag{22}$$

Notice that using τ and δ defined in Theorem 2, the bound for R_T^{SW} is $\tilde{O}(T^{\frac{2}{3}})$, even if the parameters are not optimal for this second framework.

4.2. Unknown $\{\Lambda_m\}_{m \in \mathcal{K}}$

We now focus on the situation when the auctioneer knows $\{q_i\}_{i \in \mathcal{N}}$, while $\{\Lambda_m\}_{m \in \mathcal{K}}$ are unknown. By definition of cascade model, $\{\Lambda_m\}_{m \in \mathcal{K}}$ are strictly non–increasing in m. This dramatically simplifies the allocation problem since the optimal allocation can be found without knowing the actual values of $\{\Lambda_m\}_{m \in \mathcal{K}}$. Indeed, the allocation θ^* such that $\alpha(m; \theta^*) \in \arg\max_{i \in \mathcal{N}} (q_i \hat{v}_i; m), \forall m$, is optimal for any possible $\{\Lambda_m\}_{m \in \mathcal{K}}$. However, the lack of knowledge about $\{\Lambda_m\}_{m \in \mathcal{K}}$ makes the design of a truthful mechanism non trivial because the cumulative probabilities of observation appear in the calculation of the payments. In the following, we initially focus on DSIC in expectation mechanisms, providing two mechanisms, the first DSIC in expectation w.r.t. the click realizations and the second DSIC in expectation w.r.t. the realizations of the random component of the mechanism, and finally we discuss about DSIC *a posteriori* mechanisms.

4.2.1. DSIC in expectation w.r.t. the click realizations mechanism

In this case, we do not need to estimate the parameters $\{\Lambda_m\}_{m \in \mathcal{K}}$ and therefore we do not resort to the MAB framework to solve any exploration-exploitation dilemma. The pseudocode of the algorithm A–VCG2 (Adaptive VCG2) is given in Fig. 2. On the basis of the above considerations, we can adopt the allocatively efficient allocation function f^* as prescribed by Eq. 8 even if the mechanism does not know the actual values of the parameters $\{\Lambda_m\}_{m \in \mathcal{K}}$. Nonetheless, the VCG payments defined in Eq. 9 cannot be

Input: Qualities parameters $\{q_i\}_{i\in\mathcal{N}}$

for $t = 1, \ldots, T$ **do**
  Allocate ads according to f^* as prescribed by (8)
  if Ad a_i is displayed **then**
    Ask for payment p_i^c defined in (23)

Figure 2: Pseudo–code for the A–VCG2 mechanism.

computed, since $\{\Lambda_m\}_{m\in\mathcal{K}}$ are unknown. However, by resorting to execution–contingent payments (generalizing the pay–per–click approach), we can impose computable payments that, in expectation, are equal to Eq. 9.[14] More precisely, the contingent payments are computed given the bids $\hat{\mathbf{v}}$ and all click events over the slots and take the form:

$$p_i^c\left(\hat{\mathbf{v}}, \{click_m^{\alpha(m;\theta^*)}\}_{m=1}^K\right) \tag{23}$$
$$= \sum_{\pi(i;\theta^*)\leq m\leq K} click_m^{\alpha(m;\theta^*)} \cdot \frac{q_{\alpha(m;\theta^*_{-i})} \cdot \hat{v}_{\alpha(m;\theta^*_{-i})}}{q_{\alpha(m;\theta^*)}} - \sum_{\pi(i;\theta^*)<m\leq K} click_m^{\alpha(m;\theta^*)} \cdot \hat{v}_{\alpha(m;\theta^*)}$$

Notice that the payment p_i^c depends not only on the click of ad a_i, but also on the clicks of all the ads displayed below a_i. In expectation, the two terms of p_i^c are:

$$\mathbb{E}_{click}\left[\sum_{\pi(i;\theta^*)\leq m\leq K} click_m^{\alpha(m;\theta^*)} \cdot \frac{q_{\alpha(m;\theta^*_{-i})} \cdot \hat{v}_{\alpha(m;\theta^*_{-i})}}{q_{\alpha(m;\theta^*)}}\right] = \sum_{\pi(j;\theta^*)\geq\pi(i;\theta^*)} \Lambda_{\pi(j;\theta^*_{-i})} q_j \hat{v}_j$$

$$\mathbb{E}_{click}\left[\sum_{\pi(i;\theta^*)<m\leq K} click_m^{\alpha(m;\theta^*)} \cdot \hat{v}_{\alpha(m;\theta^*)}\right] = \sum_{\pi(j;\theta^*)>\pi(i;\theta^*)} \Lambda_{\pi(j;\theta^*)} q_j \hat{v}_j$$

and therefore, in expectation, the payment equals those defined in Eq. 9. We discuss the properties of the mechanism in what follows.

Proposition 2. *The A–VCG2 is IR* a posteriori.

[14] In pay–per–click payments, an advertiser pays only once its ad is clicked; in our execution–contingent payments, an advertiser pays also when the ads of other advertisers are clicked.

PROOF. Rename the ads $\{a_1, \ldots, a_N\}$ such that $q_1 v_1 \geq q_2 v_2 \geq \ldots \geq q_N v_N$. We can write payments in Eq. 23 as:

$$p_i^c\left(\mathbf{v}, \{click_j^j\}_{j=1}^K\right) = \sum_{j=i}^{K} \frac{click_j^j}{q_j} q_{j+1} v_{j+1} - \sum_{j=i+1}^{K} click_j^j v_j$$

Thus, the utility for advertiser a_i is:

$$\begin{aligned}
u_i &= click_j^j v_i + \sum_{j=i+1}^{K} click_j^j v_j - \sum_{j=i}^{K} \frac{click_j^j}{q_j} q_{j+1} v_{j+1} \\
&= \sum_{j=i}^{K} click_j^j v_j - \sum_{j=i}^{K} \frac{click_j^j}{q_j} q_{j+1} v_{j+1} \\
&= \sum_{j=i}^{K} \left(click_j^j v_j - \frac{click_j^j}{q_j} q_{j+1} v_{j+1} \right) \\
&= \sum_{j=i}^{K} \frac{click_j^j}{q_j} (q_j v_j - q_{j+1} v_{j+1}).
\end{aligned}$$

Since $\frac{click_j^j}{q_j} \geq 0$ by definition and $q_j v_j - q_{j+1} v_{j+1} \geq 0$ because of the chosen ordering of the ads, then the utility is always positive and we can conclude that the mechanism is IR *a posteriori*. □

Theorem 4. *The A–VCG2 is:*

- *DSIC in expectation w.r.t. the click realizations,*
- *IR* a posteriori*,*
- *WBB in expectation w.r.t. the click realizations,*
- *AE.*

PROOF. It trivially follows from the fact that the allocation function is AE and the payments in expectation equal the VCG ones and by Proposition 2. □

Proposition 3. *The A–VCG2 is not DSIC* a posteriori *(w.r.t. the click realizations).*

PROOF. The proof is by counterexample. Consider an environment with 3 ads $\mathcal{N} = \{1, 2, 3\}$ and 2 slots $S = \{1, 2\}$ s.t. $q_1 = 0.5$, $v_1 = 4$, $q_2 = 1$, $v_2 = 1$, $q_3 = 1$, $v_3 = 0.5$, which correspond to expected values of 2, 1, and 0.5.

The optimal allocation θ^* consists in allocating a_1 in s_1 and a_2 in s_2. Consider a step t when both ad a_1 and a_2 are clicked, from Eq. 23 we have that the payment of a_2 is:

$$p_2^c\big(\mathbf{v}, \{click_m^{\alpha(m;\theta^*)}\}_{m=1}^K\big) = \frac{1}{q_2} q_3 v_3 = 0.5$$

If ad a_2 reports a value $\hat{v}_2 = 3$, the optimal allocation is now a_2 in s_1 e a_1 in s_2. In the case both a_1 and a_2 are clicked, the payment of a_2 is:

$$p_2^c\big(\hat{\mathbf{v}}, \{click_m^{\alpha(m;\theta^*)}\}_{j=1}^K\big) = \frac{1}{q_2} q_1 v_1 + \frac{1}{q_1} q_3 v_3 - v_1 = 2 + 1 - 4 = -1$$

Given that, in both cases, the utility is $u_2 = v_2 - p_2^c\big(\hat{\mathbf{v}}, \{click_m^{\alpha(m;\theta^*)}\}_{m=1}^K\big)$, reporting a non–truthful value is optimal. Thus, we can conclude that the mechanism is not DSIC *a posteriori* w.r.t. the click realizations. □.

Proposition 4. *The A–VCG2 is not WBB* a posteriori *(w.r.t. the click realizations).*

PROOF. The proof is by counterexample. Consider an environment with 3 ads $\mathcal{N} = \{1, 2, 3\}$ and 2 slots $S = \{1, 2\}$ s.t. $q_1 = 1$, $v_1 = 2$, $q_2 = 0.5$, $v_2 = 1$, $q_3 = 1$, $v_3 = 0.1$.

The optimal allocation θ^* consists in allocating a_1 in s_1 e a_2 in s_2. Consider step t when both ad a_1 and a_2 are clicked, their payments are:

$$p_1^c\left(\mathbf{v}, \{click_m^{\alpha(m;\theta^*)}\}_{m=1}^K\right) = \frac{1}{q_1} q_2 v_2 + \frac{1}{q_2} q_3 v_3 - v_2 = 0.5 + 2 \cdot 0.1 - 1 = 0.2 - 0.5 < 0$$

$$p_2^c = \frac{1}{q_2} q_3 v_3 = 0.2$$

Thus, $\sum_{i=1}^3 p_i^c \left(\mathbf{v}, \{click_m^{\alpha(m;\theta^*)}\}_{m=1}^K\right) = 0.4 - 0.5 < 0$, and we can conclude that the mechanism is not WBB *a posteriori*. □

Input: Qualities parameters $\{q_i\}_{i\in\mathcal{N}}$

for $t = 1, \dots, T$ **do**
Allocate ads according to $f^{*'}$ as prescribed by Algorithm 1
For each ad a_i, ask for payment $p_i^{B,*,c}$ defined in (25)

Figure 3: Pseudo–code for the A–VCG2′ mechanism.

Now we state the following theorem, whose proof is straightforward.

Theorem 5. *Let us consider a sequential auction with N advertisers, K slots, and T steps, with position–dependent cascade model with parameters $\{\Lambda_m\}_{m=1}^K$. The A–VCG2 achieves an auctioneer's revenue expected regret $R_T = 0$.*

An important property of this mechanism is that the expected payments are exactly the VCG payments for the optimal allocation when all the parameters are known. Moreover, the absence of an exploration phase allows us to obtain a per–step expected regret of zero and, thus, the cumulative regret over the T steps of auction is $R_T = 0$. Similar considerations can be applied to the study of the regret over the SW, obtaining the following.

Corollary 1. *Let us consider a sequential auction with N advertisers, K slots, and T steps, with position–dependent cascade model with parameters $\{\Lambda_m\}_{m=1}^K$. The A–VCG2 achieves an SW regret $R_T^{SW} = 0$.*

4.2.2. DSIC in expectation w.r.t. the realizations of the random component mechanisms

As for the previous mechanism, also in this case we only need an exploitation phase. Unlike A–VCG2, in this case we need to follow a similar approach as in [23] and introduce a random component, which leads to the mechanism, called A–VCG2′ reported in Fig. 3.

Since f^* is monotonic (see Appendix B) and the problem is with single parameter and linear utilities, payments that guarantee DSIC in expectation w.r.t. the click realizations can be written as Myerson payments:

$$p_i^*(\hat{\mathbf{v}}) = \Lambda_{\pi(i;f^*(\hat{\mathbf{v}}))} q_i \hat{v}_i - \int_0^{\hat{v}_i} \Lambda_{\pi(i;f^*(\hat{\mathbf{v}}_{-i},u))} q_i du, \tag{24}$$

which coincide with the VCG payments defined in Eq. 3 (hence the use of the same notation p_i^*). This is justified by the fact that when a mechanism is AE, IR in expectation w.r.t. the click realizations and WBB in expectation w.r.t. the click realizations the only payments that lead to a DSIC in expectation w.r.t. the click realizations mechanism are the VCG payments with Clacke's pivot [45], and thus Eq. 24 coincides with Eq. 3. However, these payments are not directly computable, since the parameters $\{\Lambda_m\}_{m\in\mathcal{K}}$ in the integral are unknown and, as in the case discussed in Section 4.2.1, we cannot replace them by empirical estimates. Nonetheless, we could obtain these payments in expectation by using execution–contingent payments associated with non–optimal allocations where the report $\hat{v}_i$ is randomly modified in an interval between 0 and the actual value. This can be obtained by resorting to the approach proposed in [23], which takes as input a generic allocation function f and introduces a randomized component to it, thus producing a new (random) allocation function that we denote by f'. At the cost of reducing the efficiency of the mechanism, this technique allows the computation of the allocation and the payments at the same time, even when payments described in [35] cannot be directly computed.

In A–VCG2′, we apply this approach to f^*, thus obtaining a new allocation function $f^{*'}$. With $f^{*'}$, the advertisers' reported values $\{\hat{v}_i\}_{i\in\mathcal{N}}$ are modified, each with a (small) probability $\mu \in (0, 1)$. The (potentially) modified values are then used to compute the allocation (using f^*) and the payments. More precisely, with a probability of $(1-\mu)^N$, $f^{*'}$ returns the same allocation as f^*, while it defines a different allocation with probability of $1-(1-\mu)^N$. The reported values $\{\hat{v}_i\}_{i\in\mathcal{N}}$ are modified through the canonical Self–Resampling Procedure (cSRP) described in [23] that generates two samples: $x_i(\hat{v}_i, \omega_i)$ and $y_i(\hat{v}_i, \omega_i)$, where ω_i is the random seed. We sketch the result of cSRP where the function 'rec' is defined in [23]:

$$(x_i, y_i) = cSRP(\hat{v}_i) = \begin{cases} (\hat{v}_i, \hat{v}_i) & \text{w.p. } 1-\mu \\ (\hat{v}_i'', \hat{v}_i') & \text{otherwise} \end{cases},$$

where $\hat{v}_i' \sim \mathcal{U}([0, \hat{v}_i])$ and $\hat{v}_i'' = \text{rec}(\hat{v}_i')$. The algorithm in Fig. 4 shows how $f^{*'}$ works when the original allocation function is f^*. The reported values $\{\hat{v}_i\}_{i\in\mathcal{N}}$ are perturbed through the cSRP (Step 2) and then the allocation is chosen by applying the original allocation function f^* to the new values $\mathbf{x}$ (Step 4). Finally, the payments are computed as

for all a_i, $i \in \mathcal{N}$ **do**
  $(x_i, y_i) = cSRP(\hat{v}_i)$
  $\mathbf{x} = (x_1, \ldots, x_N)$
$\theta = f^*(\mathbf{x})$

Figure 4: Definition of $f^{*'}(\hat{\mathbf{v}})$.

$$p_i^{B,*,c}(\mathbf{x}, click^i_{\pi(i;f^*(\mathbf{x}))}) = \begin{cases} \frac{p_i^{B,*}(\mathbf{x},\mathbf{y};\hat{\mathbf{v}})}{\Lambda_{\pi(i;f^*(\mathbf{x}))}q_i} & \text{if } click^i_{\pi(i;f^*(\mathbf{x}))} = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \hat{v}_i - \begin{cases} \frac{1}{\mu}\hat{v}_i & \text{if } y_i < \hat{v}_i \\ 0 & \text{otherwise,} \end{cases} & \text{if } click^i_{\pi(i;f^*(\mathbf{x}))} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

where

$$p_i^{B,*}(\mathbf{x}, \mathbf{y}; \hat{\mathbf{v}}) = \Lambda_{\pi(i;f^*(\mathbf{x}))}q_i\hat{v}_i - \begin{cases} \frac{1}{\mu}\Lambda_{\pi(i;f^*(\mathbf{x}))}q_i\hat{v}_i & \text{if } y_i < \hat{v}_i \\ 0 & \text{otherwise} \end{cases}, \tag{26}$$

with $\mathbf{y} = (y_1, \ldots, y_N)$. If we take the expectation of the payments in Eq. 25 w.r.t. the randomization of the mechanism, then we obtain exactly the same form of payments as in Eq. 24 but instantiated for the randomized allocation function $f^{*'}$ (for the explicit equation refer to Eq. E.1 in Appendix E). Furthermore, the resulting mechanism is shown to be DSIC in expectation w.r.t. the realizations of the random component of the mechanism and *a posteriori* w.r.t. the click realizations. As a result we obtained the following properties.

Theorem 6. *The A–VCG2 ′ is:*

- *DSIC in expectation w.r.t. the realizations of the random component of the mechanism and* a posteriori *w.r.t. the click realizations,*
- *IR* a posteriori*,*
- *WBB in expectation w.r.t. the realizations of the random component of the mechanism and* a posteriori *w.r.t. the click realizations.*

We also obtain the following regret guarantees.

Theorem 7. *Let us consider an auction with N advertisers, K slots, and T steps, with position–dependent cascade model with parameters $\{\Lambda_m\}_{m=1}^K$. The A–VCG2′ achieves an auctioneer's revenue expected regret $R_T \leq 2K^2\mu v_{\max}T$.*

If we tune the randomization parameter as $\mu = \frac{1}{K^2T}$, then we obtain $R_T = O(1)$. Notice that μ cannot be just set to zero, since it would lead to a division by zero in Eq. 25. Furthermore, as illustrated in Section 6, an undesirable effect of a small μ is the corresponding increase in the variance in the payments. Thus a proper trade-off should be found when tuning μ in practice. We provide a similar result for the regret over the SW.

Theorem 8. *Let us consider a sequential auction with N advertisers, K slots, and T steps, with position–dependent cascade model with parameters $\{\Lambda_m\}_{m=1}^K$. The A–VCG2′ achieves a SW regret $R_T^{SW} \leq K^2\mu v_{\max}T$.*

4.2.3. Discussion about DSIC a posteriori *mechanisms*

One may wonder whether there exists a no–regret DSIC *a posteriori* mechanism, even at the cost of a worse regret. Resorting to the same arguments used in [28], we show that the answer to such question is negative.

Theorem 9. *Let us consider a sequential auction with N advertisers, K slots, and T steps, with position–dependent cascade model with parameters $\{\Lambda_m\}_{m=1}^K$ whose value are unknown. Any online learning mechanism that is DSIC* a posteriori *achieves an auctioneer's revenue expected regret R_T of $\Theta(T)$.*

Proof. **(sketch)** Basically, the A–VCG2 mechanism is DSIC in expectation w.r.t. the click realizations because it adopts execution–contingent payments in which the payment of advertiser a_i depends also on the clicks over ads other than a_i, while A–VCG2′ is DSIC in expectation w.r.t. the realization of the random component of the mechanism because it adopts implicit payments. In order to have DSIC *a posteriori*, we need payments p_i that are deterministic w.r.t. the click realizations over other ads other than a_i (i.e., pay–per–click payments are needed) and deterministic w.r.t. any realization of the random component of the mechanism.

We notice that even if $\{\Lambda_m\}_{m\in\mathcal{K}}$ have been estimated (e.g., in an exploitation phase), we cannot have payments leading to DSIC *a posteriori*. Indeed, with estimates $\{\tilde{\Lambda}_m\}_{m\in\mathcal{K}}$, the allocation function maximizing $\widetilde{SW}$ (computed

with $\tilde{\Lambda}_m$) is not an affine maximizer and therefore the adoption of WVCG mechanism would not guarantee DSIC, not even in expectation. As a result, only mechanisms with payments defined in Eq. 24 can be used. However, these payments, if computed exactly (and not estimated in expectation), as required to have DSIC *a posteriori*, require the knowledge about the actual Λ_m related to each slot s_m in which an ad can be allocated for each report $\hat{v} \leq v$.

To prove the theorem, we provide a characterization of DSIC *a posteriori* mechanisms. More precisely, we need a monotonic allocation function and the payments defined in Eq. 24. As mentioned above, these payments require the knowledge about the actual Λ_m related to the slot s_m in which an ad can be allocated for any report $\hat{v} \leq v$. Thus we have two possibilities:

- In the first case, the ads are partitioned and each partition is associated with a single slot and the ad with the largest expected valuation is chosen at each slot independently. In other words, an ad can be allocated only in one given specific slot and its report determines only whether it is displayed or not (and not where). This case is equivalent to multiple separate–single slot auctions and therefore each auction is DSIC *a posteriori* as shown in [20]. However, as shown in [28], this mechanism would have a regret $\Theta(T)$.

- In the second case, the ads are partitioned and each partition is associated with multiple slots and for each partition an auction is carried out to determine the allocation over each slot. In other words, an ad can be allocated in one of a given set of slots (associated with its partition) on the basis of its report. In this case, to compute the payments, it would be necessary to know the exact CTR of the ad for each possible slot, but this is possible only in expectation either by using the above execution–contingent, as we do in Section 4.2.1, or by using a random component in the mechanism, as we do in Section 4.2.2. However, in both these case we would not obtain DSIC *a posteriori*.

Thus, in order to have DSIC *a posteriori*, we need to adopt the class of mechanisms described in the first case, obtaining $R_T = \Theta(T)$. □

4.3. Unknown $\{\Lambda_m\}_{m\in\mathcal{K}}$ and $\{q_i\}_{i\in\mathcal{N}}$

In this section we study the situation in which both $\{q_i\}_{i\in\mathcal{N}}$ and $\{\Lambda_m\}_{m\in\mathcal{K}}$ are unknown. From the results derived in the previous section, we know that

Input: Length of exploration phase τ, confidence δ

Exploration phase
for $t = 1, \dots, \tau$ **do**
  Allocate ads according to (13)
  Ask for no payment
  Observe the clicks $\{click_1^i(t)\}_{i=1}^N$
Compute the estimated quality $\tilde{q}_i = \frac{1}{|B_i|} \sum_{t \in B_i} click_1^i(t)$
Compute $\tilde{q}_i^+ = \tilde{q}_i + \eta$ where η is given by (27)

Exploitation phase
for $t = \tau + 1, \dots, T$ **do**
  Allocate ads according to $\tilde{f}'$ as prescribed by Algorithm 1 adopting $\tilde{f}$ instead of f^*
  For each ad a_i, ask for payment $\tilde{p}_i^{B,c}$ defined in (28)

Figure 5: Pseudo–code for the A–VCG3 mechanism.

adopting DSIC *a posteriori* leads to $R_T = \Theta(T)$. Thus, we will only focus on DSIC in expectation.

First of all, we remark that the mechanisms presented in Section 4.1 and 4.2 cannot be adopted here and a new mechanism is needed. By combining A–VCG1 and A–VCG2′, we obtain the algorithm A–VCG3 (Adaptive VCG3) illustrated in Fig. 5. As in the case when only the qualities $\{q_i\}_{i \in \mathcal{N}}$ are unknown, we formalize the problem as a MAB where the exploration and exploitation phases are separate and where, during the exploration phase, we estimate the values of $\{q_i\}_{i \in \mathcal{N}}$.

Exploration phase. During the first τ steps, estimates of $\{q_i\}_{i \in \mathcal{N}}$ are computed. We use the same exploration policy of Section 4.1, but the estimates are computed just using samples from the first slot, since Λ_m with $m > 1$ are unknown.[15] Define $B_i = \{t : \pi(i; \theta_t) = 1, 1 \leq t \leq \tau\}$ the set of steps

[15]In the following, we report some considerations about the case in which also the samples from the slots below the first are considered. Let us observe that, even if we use only the samples from the first slot, the algorithms [20, 21] that apply to the single–slot case cannot be adopted here unless to accept a regret $\Theta(T)$. This is essentially due to the fact that algorithms [20, 21] have deterministic payments, but, as we show in Section 4.2.3,

where a_i is displayed in the first slot, the number of samples collected for a_i is $|B_i| = \lfloor \frac{\tau}{N} \rfloor \geq \frac{\tau}{2N}$.[16] The estimated value of q_i is:

$$\tilde{q}_i = \frac{1}{|B_i|} \sum_{t \in B_i} click_1^i(t).$$

such that $\tilde{q}_i$ is an unbiased estimate of q_i (i.e., $\mathbb{E}_{click}[\tilde{q}_i] = q_i$). By applying the Hoeffding's inequality we obtain an upper bound over the error of the estimated quality $\tilde{q}_i$ for each ad a_i.

Proposition 5. *For any ad* $\{a_i\}_{i \in \mathcal{N}}$

$$|q_i - \tilde{q}_i| \leq \sqrt{\frac{1}{2|B_i|} \log \frac{2N}{\delta}} \leq \sqrt{\frac{N}{\tau} \log \frac{2N}{\delta}} =: \eta, \tag{27}$$

with probability $1 - \delta$ *(w.r.t. the click realizations).*

In this case, in order to have a meaningful bound, i.e., $|q_i - \tilde{q}_i| < 1$, the length of the exploration phase has to be $\tau > N \log \frac{2N}{\delta}$. After the exploration phase, an upper–confidence bound over each quality is computed as $\tilde{q}_i^+ = \tilde{q}_i + \eta$.

Exploitation phase. We first focus on the allocation function. During the exploitation phase we want to use an allocation $\tilde{\theta} = \tilde{f}(\hat{\mathbf{v}})$ maximizing the estimated SW with estimated $\{\tilde{q}_i^+\}_{i \in \mathcal{N}}$ and the parameters $\{\Lambda_m\}_{m \in \mathcal{K}}$. Since the actual parameters $\{\Lambda_m\}_{m \in \mathcal{K}}$ are monotonically non–increasing, $\tilde{\theta}$ is defined as an allocation $\{\langle s_m, a_{\alpha(m;\tilde{\theta})} \rangle\}_{m \in \mathcal{K}'}$, where

$$\alpha(m; \tilde{\theta}) \in \arg\max_{i \in \mathcal{N}} (\tilde{q}_i^+ \hat{v}_i; m) = \arg\max_{i \in \mathcal{N}} (\tilde{q}_i^+ \Lambda_m \hat{v}_i; m).$$

We now focus on payments. Allocation function $\tilde{f}$ is an affine maximizer (due to weights depending on $\tilde{q}_i$ as in Section 4.1), but WVCG payments cannot be computed given that parameters $\{\Lambda_m\}_{m \in \mathcal{K}}$ are unknown. Neither the adoption of execution–contingent payments, like in Eq. 23, is allowed,

we cannot have no–regret mechanisms when payments are deterministic.

[16] Following the same reasoning of Section 4.1, we consider an exploration time of $\tau > 2N$, which guarantees to have at least two samples to estimate each $\tilde{q}_i^+$.

given that q_i is unknown and only estimates $\tilde{q}_i$ are available. Thus, we resort to implicit payments as in Section 4.2.2. More precisely, we use the same exploitation phase adopted in Section 4.2.2, except that we use $\tilde{f}$ in place of f^*. In this case, we have that the per–click payments are:

$$\tilde{p}_i^{B,c}(\mathbf{x}, click^i_{\pi(i;\tilde{f}(\mathbf{x}))}) = \begin{cases} \frac{\tilde{p}_i^B(\mathbf{x},\mathbf{y};\hat{\mathbf{v}})}{\Lambda_{\pi(i;\tilde{f}(\mathbf{x}))} q_i} & \text{if } click^i_{\pi(i;\tilde{f}(\mathbf{x}))} = 1 \\ 0 & \text{otherwise} \end{cases} =$$

$$= \begin{cases} \hat{v}_i - \begin{cases} \frac{1}{\mu}\hat{v}_i & \text{if } y_i < \hat{v}_i \\ 0 & \text{otherwise} \end{cases} & \text{if } click^i_{\pi(i;\tilde{f}(\mathbf{x}))} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

where

$$\tilde{p}_i^B(\mathbf{x}, \mathbf{y}; \hat{\mathbf{v}}) = \Lambda_{\pi(i;\tilde{f}(\mathbf{x}))} q_i \hat{v}_i - \begin{cases} \frac{1}{\mu}\Lambda_{\pi(i;\tilde{f}(\mathbf{x}))} q_i \hat{v}_i & \text{if } y_i < \hat{v}_i \\ 0 & \text{otherwise} \end{cases}. \tag{29}$$

We can state the following.

Theorem 10. *The A–VCG3 is:*

- *DSIC in expectation w.r.t. the realizations of the random component of the mechanism and* a posteriori *w.r.t. the click realizations,*
- *IR* a posteriori*,*
- *WBB in expectation w.r.t. the realizations of the random component of the mechanism and* a posteriori *w.r.t. the click realizations.*

Proof. The proof of DSIC in expectation and WBB in expectation easily follows from the definition of the adopted mechanism as discussed in [23]. The proof of IR *a posteriori* is similar to the proof of Proposition 2. The fact that the properties hold *a posteriori* w.r.t. the click realizations follows from [23]. □

Now we want to analyze the performance of the mechanism in terms of the regret accumulated through T steps. Notice that in this case we have to focus on two different potential sources of regret: the adoption of a sub–optimal (randomized) allocation function and the estimation of the unknown parameters.

Theorem 11. *Let us consider a sequential auction with N advertisers, K slots, and T steps, with position–dependent cascade model with parameters $\{\Lambda_m\}_{m=1}^K$, accuracy η as defined in Eq. 27 and parameter $\mu \in (0,1]$. For any parameter $\tau \in \{1,\dots,T\}$ and $\delta \in (0,1)$, the A–VCG3 achieves an auctioneer's revenue expected regret:*

$$R_T \leq v_{\max} K \left[(T-\tau)\left(2\eta + 2\mu N\right) + \tau + \delta T \right].$$

By setting the parameters to

- $\mu = T^{-\frac{1}{3}} N^{-\frac{2}{3}}$,
- $\delta = T^{-\frac{1}{3}} N^{\frac{1}{3}}$,
- $\tau = T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log \frac{2N}{\delta}\right)^{\frac{1}{3}}$,

then the regret is

$$R_T \leq 6 v_{\max} K T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log \left(2 T^{\frac{1}{3}} N^{\frac{2}{3}} \right) \right)^{\frac{1}{3}}. \tag{30}$$

Remark 1 (The bound). Up to numerical constants and logarithmic factors, the previous bound on $R_T = \tilde{O}(T^{\frac{2}{3}} K N^{\frac{1}{3}})$. We first notice that also in this case we match the lowest possible regret w.r.t. T when exploration and exploitation phases are separate. As a result, the proposed mechanism is a no–regret algorithm and it asymptotically approaches the performance of the VCG (when all the parameter are known). Compared to the results in Section 4.1, the dependency of the regret on K increased by a factor $K^{\frac{1}{3}}$ and it is now linear. This is a direct consequence of the exploration phase. In fact, here we cannot take advantage of the samples collected over all the slots, and the qualities are estimated only using samples observed in the first slot. On the other hand, the dependency on N is the same as in Section 4.1.

Remark 2 (Non–separate phases and $\tilde{O}(T^{\frac{1}{2}})$). The questions whether it is possible to avoid separating exploration and exploitation and preserve DSIC in expectation (in some form) and whether it is possible to obtain a regret of $\tilde{O}(T^{\frac{1}{2}})$ are open.

Remark 3 (Using samples from multiple slots). An important issue is whether it is possible to exploit samples from the slots below the first one to improve the accuracy of the estimates and reduce the length of the exploration phase. The critical issue here is that the samples from slots below

the first are drawn from a Bernoulli distribution with parameter obtained by the product of Λ_m and q_i, and it is not trivial to find a method to use these samples to improve the estimates. However, we notice that the exploitation of these additional samples would correspond to a reduction of the regret bound of at most $K^{\frac{1}{3}}$, given that the dependency from K cannot be better than in the case discussed in Section 4.1 (i.e., $O(K^{\frac{2}{3}})$).

We can also prove an upper–bound for the regret for the SW of A–VCG3.

Theorem 12. *Let us consider an auction with N advertisers, K slots, and T steps, with position–dependent cascade model with parameters $\{\Lambda_m\}_{m=1}^{K}$, accuracy η as defined in Eq. 27 and parameter $\mu \in (0, 1]$. For any parameter $\tau \in \{1, \dots, T\}$ and $\delta \in (0, 1)$, the A–VCG3 achieves a SW regret:*

$$R_T^{SW} \leq v_{\max} K \left[(T - \tau)(2\eta + N\mu) + \tau + \delta T \right].$$

By setting the parameters to

- $\mu = K^{-1} T^{-\frac{1}{3}} N^{\frac{1}{3}}$,
- $\delta = T^{-\frac{1}{3}} N^{\frac{1}{3}}$,
- $\tau = T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log \frac{2N}{\delta} \right)^{\frac{1}{3}}$,

then the regret is

$$R_T^{SW} \leq 5 v_{\max} K T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log 2 T^{\frac{1}{3}} N^{\frac{2}{3}} \right)^{\frac{1}{3}}.$$

Also in this case we obtain a regret on the SW $R_T^{SW} = \tilde{O}(T^{\frac{2}{3}})$.

5. Learning with Position– and Ad–Dependent Externalities

In this section we deal with the general model in Eq. 1, where both position– and ad–dependent externalities are present and we provide several partial results. In Section 5.1, we analyze the problem of designing a DSIC *a posteriori* mechanism when only the qualities of the ads are unknown, while in Section 5.2 we highlight some problems that rise when also continuation probabilities are uncertain.

Input: Length of exploration phase τ, confidence δ, position–dependent parameters $\{\Gamma_m\}_{m\in\mathcal{K}}$

Exploration phase
for $t = 1, \dots, \tau$ **do**
  Allocate ads according to (13)
  Ask for no payment
  Observe the clicks $\{click^i_{\pi(i;\theta_t)}(t)\}_{i=1}^N$
Compute the estimated quality $\tilde{q}_i = \frac{1}{|B_i|}\sum_{t\in B_i} \frac{click^i_{\pi(i;\theta_t)}(t)}{\Gamma_{\pi(i;\theta_t)(\theta_t)}}$
Compute $\tilde{q}_i^+ = \tilde{q}_i + \eta$ where η is given by (31)

Exploitation phase
for $t = \tau+1, \dots, T$ **do**
  Allocate ads according to $\tilde{f}$
  if Ad a_i is clicked **then**
    Ask for payment $\tilde{p}_i^c$ defined in (32)

Figure 6: Pseudo–code for the A–VCG4 mechanism.

5.1. Unknown qualities $\{q_i\}_{i\in\mathcal{N}}$

We first focus on the problem where the only unknown parameters are the qualities $\{q_i\}_{i\in\mathcal{N}}$ of the ads and the externality model includes position– and ad–dependent externalities. We focus on DSIC in expectation w.r.t. the click realizations, since there is not any no–regret mechanism that is DSIC *a posteriori* w.r.t. the click realizations [43], and we study MAB algorithms that separate the exploration and exploitation phases. The structure of the mechanism we propose, called A–VCG4, is similar to the A–VCG1 and is reported in Fig. 6.

Exploration phase. At each step of the exploration phase of length τ, we collect K samples of no–click/click events. Given a generic exploration policy $\{\theta_t\}_{t=1}^\tau$, the estimated quality $\tilde{q}_i$ is computed as:

$$\tilde{q}_i = \frac{1}{|B_i|}\sum_{t\in B_i} \frac{click^i_{\pi(i;\theta_t)}(t)}{\Gamma_{\pi(i;\theta_t)}(\theta_t)},$$

where $B_i = \{t : \pi(i;\theta_t) \le K, 1 \le t \le \tau\}$. Since the explorative allocations θ_t impact on the cumulative probabilities of observation $\Gamma_m(\theta_t)$, we use a

variation of Proposition 1 in which Eq. 12 is replaced by:

$$|q_i - \tilde{q}_i| \leq \sqrt{\left(\sum_{t \in B_i} \frac{1}{\Gamma_{\pi(i;\theta_t)}(\theta_t)^2}\right) \frac{1}{2|B_i|^2} \log \frac{2N}{\delta}}.$$

For each exploration policy such that $|B_i| = \lfloor K\tau/N \rfloor \geq \frac{K\tau}{2N}$ for any $i \in \mathcal{N}$ (e.g., the policy defined in Eq. 13), we redefine η as

$$|q_i - \tilde{q}_i| \leq \frac{1}{\Gamma_{\min}} \sqrt{\frac{N}{K\tau} \log \frac{2N}{\delta}} =: \eta, \tag{31}$$

where $\Gamma_{\min} = \min_{\theta \in \Theta, m \in \mathcal{K}} \{\Gamma_m(\theta)\}$. In this case, in order to have a meaningful bound, i.e., $|q_i - \tilde{q}_i| < 1$, the length of the exploration phase has to be $\tau > \frac{1}{\Gamma^2_{\min}} \frac{N}{K} \log \frac{2N}{\delta}$. We define the upper–confidence bound $\tilde{q}_i^+ = \tilde{q}_i + \eta$. During the exploration phase, in order to preserve the DSIC *a posteriori* property, the allocations $\{\theta_t\}_{t=1}^{\tau}$ do not depend on the reported values of the advertisers and no payments are imposed to the advertisers.

Exploitation phase. We define an upper bound on the SW as

$$\widetilde{\mathrm{SW}}(\theta, \hat{\mathbf{v}}) = \sum_{i=1}^{N} \Gamma_{\pi(i;\theta)}(\theta) \tilde{q}_i^+ \hat{v}_i = \sum_{m=1}^{K} \Gamma_m(\theta) \tilde{q}^+_{\alpha(m;\theta)} \hat{v}_{\alpha(m;\theta)}.$$

We denote by $\tilde{\theta}$ the allocation maximizing $\widetilde{\mathrm{SW}}(f(\hat{\mathbf{v}}), \hat{\mathbf{v}})$ and by $\tilde{f}$ the allocation function returning $\tilde{\theta}$:

$$\tilde{\theta} = \tilde{f}(\hat{\mathbf{v}}) \in \arg\max_{\theta \in \Theta} \widetilde{\mathrm{SW}}(\theta, \hat{\mathbf{v}}).$$

Once the exploration phase is over, the ads are allocated on the basis of $\tilde{f}$. Since $\tilde{f}$ is an affine maximizer, the mechanism can impose WVCG payments to the advertisers satisfying the DSIC *a posteriori* property. In a *pay–per–click* fashion the advertiser a_i is charged

$$\tilde{p}_i^c(\hat{\mathbf{v}}, click^i_{\pi(i;\tilde{\theta})}) = \frac{\widetilde{\mathrm{SW}}(\tilde{\theta}_{-i}) - \widetilde{\mathrm{SW}}_{-i}(\tilde{\theta})}{\Gamma_{\pi(i;\tilde{\theta})}(\tilde{\theta}) \tilde{q}_i^+} click^i_{\pi(i;\tilde{\theta})}, \tag{32}$$

which corresponds, in expectation, to the WVCG payment $\mathbb{E}\left[\tilde{p}_i^c(\hat{\mathbf{v}}, click^i_{\pi(i;\tilde{\theta})})\right] = \tilde{p}_i(\hat{\mathbf{v}})$. As a result, we have:

Theorem 13. *The A–VCG4 is:*

- *DSIC in expectation w.r.t. the click realizations,*
- *IR* a posteriori*,*
- *WBB* a posteriori*.*

We are interested in bounding the regret of the auctioneer's revenue due to A–VCG4 compared to the auctioneer's revenue of the VCG mechanism when all the parameters are known.

Theorem 14. *Let us consider a sequential auction with N advertisers, K slots, and T steps with position/ad–dependent externalities and cumulative discount factors $\{\Gamma_m(\theta)\}_{m=1}^K$ and accuracy η defined as in Eq. 31. For any parameter $\tau \in \{1, \dots, T\}$ and $\delta \in (0,1)$, the A–VCG4 achieves an auctioneer's revenue expected regret:*

$$R_T \leq v_{\max} K \left[\frac{10K}{q_{\min}} (T - \tau)\eta + \tau + \delta T \right], \tag{33}$$

where $q_{\min} = \min_{i \in \mathcal{N}} q_i$. By setting the parameters to

- $\delta = \left(\frac{5}{\Gamma_{\min}}\right)^{\frac{2}{3}} K^{\frac{1}{3}} T^{-\frac{1}{3}} N^{\frac{1}{3}}$,
- $\tau = \left(\frac{5}{\Gamma_{\min}}\right)^{\frac{2}{3}} K^{\frac{1}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log \frac{2N}{\delta}\right)^{\frac{1}{3}}$,

then the regret is

$$R_T \leq \frac{4 \cdot 5^{\frac{2}{3}}}{\Gamma_{\min}^{\frac{2}{3}} q_{\min}} v_{\max} K^{\frac{4}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log \frac{2\Gamma_{\min}^{\frac{2}{3}} K^{-\frac{1}{3}} T^{\frac{1}{3}} N^{\frac{2}{3}}}{5^{\frac{2}{3}}} \right)^{\frac{1}{3}}. \tag{34}$$

Remark 1 (Differences w.r.t. position–dependent externalities.) Up to constants and logarithmic factors, the previous distribution–free bound on R_T is $\tilde{O}(T^{\frac{2}{3}} N^{\frac{1}{3}} K^{\frac{4}{3}})$.[17] We first notice that moving from position– to

[17] We notice that in [29] we provided a bound $O(T^{\frac{2}{3}} N K^{\frac{2}{3}})$ that did not match with the numerical simulations and we conjectured a bound of $O(T^{\frac{2}{3}} N^{\frac{1}{3}} K^{\frac{4}{3}})$. Here we show the conjecture is actually correct.

position/ad–dependent externalities does not change the dependency of the regret on the number of steps T and the number of ads N. Moreover, the per–step regret still decreases to 0 as T increases. The main difference w.r.t. the bound in Theorem 2 is in the dependency on K and on the smallest quality $q_{\min}$. We believe that the augmented dependence in K is mostly due to an intrinsic difficulty of the position/ad–dependent externalities. As a result, the bound displays a super–linear dependency on the number of slots. The other main difference is that now the regret has an inverse dependency on the smallest quality $q_{\min}$. Inspecting the proof, this dependency is due to the fact that the error of a quality estimate for an ad a_i might be amplified by the inverse of the quality itself. As discussed in Remark 2 of Theorem 2, this dependency may also follow from the fact that we have a distribution–free bound. Further discussion on the tightness of this bound is postponed to Section 6.

Remark 2 (Optimization of the parameter τ). Although the actual qualities $\{q_i\}_{i\in\mathcal{N}}$ are unknown, whenever a lower-bound on $q_{\min}$ is available, the parameter τ could be better tuned by multiplying it by $(q_{\min})^{-\frac{2}{3}}$, thus reducing the regret from $\tilde{O}((q_{\min})^{-1})$ to $\tilde{O}((q_{\min})^{-\frac{2}{3}})$.

Remark 3 (Externalities–dependent bound). We notice that the previous bound does not reduce to the bound in Eq. 20 in which only position–dependent externalities are present. Indeed, the dependency on K is different in the two bounds: from $K^{\frac{2}{3}}$ in Eq. 20 to $K^{\frac{4}{3}}$ in Eq. 34. This means that the bound in Eq. 34 over–estimates the dependency on K whenever the auction has only position–dependent externalities. It is an interesting open question whether it is possible to derive an *auction–dependent* bound where the specific values of the cumulative probabilities of observation $\gamma_{m,i}$ explicitly appear in the bound and which reduces to Eq. 20 for position–dependent externalities.

(Comment to the proof). While the proof of Theorem 2 could exploit the specific definition of the payments for position–dependent slots and it is a fairly simple extension of [20], in this case the proof is more complicated because of the dependency of the cumulative probabilities of observation on the actual allocations and decomposes the regret of the exploitation phase in components due to the different allocations ($\tilde{f}$ instead of f^*) and the different qualities as well ($\tilde{q}_i^+$ instead of q_i).

Using the mechanism described before, it is possible to derive an upper–bound over the cumulative regret over the SW of the allocation (as in [23]).

We obtain the same dependence over T, as for the regret on the revenue.

Theorem 15. *Let us consider a sequential auction with N advertisers, K slots, and T steps. The auction has position/ad–dependent externalities and cumulative discount factors $\{\Gamma_m(\theta)\}_{m=1}^K$ and accuracy η defined as in Eq. 31. For any parameter $\tau \in \{1, \dots, T\}$ and $\delta \in (0,1)$, the A–VCG4 achieves a SW regret:*

$$R_T^{SW} \leq v_{\max} K \left[2(T-\tau)\eta + \tau + \delta T\right]. \tag{35}$$

By setting the parameters to

- $\delta = \left(\frac{1}{\Gamma_{\min}}\right)^{\frac{2}{3}} K^{-\frac{1}{3}} T^{-\frac{1}{3}} N^{\frac{1}{3}}$,
- $\tau = \left(\frac{1}{\Gamma_{\min}}\right)^{\frac{2}{3}} K^{-\frac{1}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log \frac{2N}{\delta}\right)^{\frac{1}{3}}$,

then the regret is

$$R_T^{SW} \leq 4 v_{\max} \left(\frac{1}{\Gamma_{\min}}\right)^{\frac{2}{3}} K^{\frac{2}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log 2\Gamma_{\min}^{-\frac{2}{3}} K^{\frac{1}{3}} T^{\frac{1}{3}} N^{\frac{2}{3}}\right)^{\frac{1}{3}}. \tag{36}$$

Thus $R_T^{SW} = \tilde{O}(T^{\frac{2}{3}})$. In particular notice that A–VCG4 is a zero–regret algorithm. We notice that unlike the bound on the revenue regret, in this case R_T^{SW} does not display any dependency on $q_{\min}$, which suggests that the problem of minimizing the SW regret may be easier. Roughly speaking, this is due to the fact that the accuracy of the estimated qualities is only used to determine the allocation $\tilde{f}$ but they do not determine the performance of $\tilde{f}$ itself, which is measured according to its actual SW. On the other hand, in the computation of the revenue regret, the qualities $\tilde{q}_i^+$ are used to determine the payments and this may lead to an additional error, which is reflected in the presence of $q_{\min}$ in the bound.

5.2. Further extensions

In this section we provide a negative, in terms of regret, result under DSIC in expectation w.r.t. the click realizations and *a posteriori* w.r.t. the realization of the random component of the mechanism when the parameter $\gamma_{m,i}$ depends only on the ad a_i (we denote it by $c_i = \gamma_{m,i}$ for any $m \in \mathcal{K}$ as

in [24]) and this parameter is the only uncertain parameter (i.e., the qualities are known).

We focus on the exploitation phase, supposing the exploration phase has produced the estimates $\{\tilde{c}_i^+\}_{i\in\mathcal{N}}$ for the continuation probabilities $\{c_i\}_{i\in\mathcal{N}}$. The allocation function f presented in [24] is able to compute the optimal allocation when $\{c_i\}_{i\in\mathcal{N}}$ values are known, but it is not an affine maximizer when applied to the estimated values $\{\tilde{c}_i^+\}_{i\in\mathcal{N}}$. In fact, in that case the allocation becomes

$$\tilde{f}(\hat{\mathbf{v}}) \in \arg\max_{\theta\in\Theta} \sum_{m=1}^{K} q_{\alpha(m;\theta)}\hat{v}_{\alpha(m;\theta)} \prod_{h=1}^{m-1} \tilde{c}^+_{\alpha(h;\theta)}. \tag{37}$$

In this case, it is not possible to isolate a weight depending only on a single ad and thus $\tilde{f}(\hat{\mathbf{v}})$ is not affine. Furthermore, we can also show that such allocation function is not monotonic.

Proposition 6. *The allocation function $\tilde{f}$ in Eq. 37 is not monotonic.*

Proof. The proof is by counterexample. Consider an environment with 3 ads and 2 slots such that

ad	v_i	$\tilde{c}_i^+$	c_i
a_1	0.85	1	0.89
a_2	1	0.9	0.9
a_3	1.4	0	0

and $q_i = 1\ \forall i \in \mathcal{N}$. The optimal allocation $\tilde{\theta}$ computed by $\tilde{f}$ when agents declare their true values $\mathbf{v}$ is: ad a_2 is allocated in the first slot and a_3 in the second one. We have $CTR_{a_3}(\tilde{\theta}) = 0.9$.

If advertiser a_3 reports a larger value, e.g., $\hat{v}_3 = 1.6$, in the resulting allocation $\tilde{f}(\hat{v}_3, \mathbf{v}_{-3})$, ad a_1 is displayed into the first slot and a_3 into the second one. In this case $CTR_{a_3}(\hat{\theta}) = 0.89 < CTR_{a_3}(\tilde{\theta})$, thus the allocation function $\tilde{f}$ is not monotonic. □

On the basis of this result, we can state the following theorem.

Theorem 16. *Let us consider a sequential auction with N advertisers, K slots, and T steps, with ad–dependent cascade model with parameters $\{c_i\}_{i=1}^N$ whose value are unknown. Any online learning mechanism that is DSIC in expectation w.r.t. the click realizations and* a posteriori *w.r.t. the realization of the random component of the mechanism achieves a SW regret $R_T^{SW} = \Theta(T)$.*

PROOF. Let, with abuse of notation, $f(\hat{\mathbf{v}}|\mathbf{c})$ be the allocation function maximizing the SW given parameters $\mathbf{c}$. As shown above, $f(\hat{\mathbf{v}}|\tilde{\mathbf{c}})$ cannot be used in the exploitation phase, because the resulting mechanism would not be DSIC in expectation w.r.t. the click realizations. However, it can be easily observed that a necessary condition to have a no–regret algorithm is that the allocation function used in the exploitation phase, say $g(\hat{\mathbf{v}}|\tilde{\mathbf{c}})$, is such that $g(\hat{\mathbf{v}}|\mathbf{c}) = f(\hat{\mathbf{v}}|\mathbf{c})$ for every $\hat{\mathbf{v}}$ and $\mathbf{c}$ (that is, they always return the same allocation). Indeed, if there exists at least a $\hat{\mathbf{v}}$ such that $g(\hat{\mathbf{v}}|\mathbf{c}) \neq f(\hat{\mathbf{v}}|\mathbf{c})$, then, as $T \to +\infty$, $f(\hat{\mathbf{v}}|\mathbf{c}) \neq g(\hat{\mathbf{v}}|\tilde{\mathbf{c}})$, given that $\tilde{\mathbf{c}} \to \mathbf{c}$. Thus, since the difference between the values of the allocations is generically strictly positive, the algorithm would suffer from a strictly positive regret when $T \to +\infty$ and therefore it would not be a no–regret mechanism. However, any such a g would not be monotonic and therefore it cannot be adopted in a DSIC in expectation w.r.t. the click realizations mechanism. As a result, any online learning mechanism that is DSIC in expectation w.r.t. the click realizations is not a no–regret mechanism.

To complete the proof, we need to provide a mechanism with regret $\Theta(T)$. Such a mechanism can be easily obtained by partitioning ads in groups such that in each group the ads compete only for a single slot. Therefore, each ad can appear in only one slot. □

The above result shows that no approach similar to the one described in [23] can be adopted even for obtaining DSIC in expectation w.r.t. realizations of the random component of the mechanism. Indeed, the approach described in [23] requires in input a monotonic allocation function. This would suggest a negative result in terms of regret also with DSIC in expectation w.r.t. realizations of the random component of the mechanism.

Finally, we provide a result on the regret over the auctioneer's revenue. The proof is straightforward given that the WVCG cannot be adopted due to the above result and therefore the regret over the payments cannot go to zero as T goes to infinite.

Theorem 17. *Let us consider a sequential auction with N advertisers, K slots, and T steps, with ad–dependent cascade model with parameters $\{c_i\}_{i=1}^N$ whose value are unknown. Any online learning mechanism that is DSIC in expectation w.r.t. the click realizations and* a posteriori *w.r.t. realization of the random component of the mechanism achieves an auctioneer's revenue expected regret $R_T = \Theta(T)$.*

6. Numerical Simulations

In this section we report numerical simulations to validate the theoretical bounds over the regret of the auctioneer's revenue proved in the previous sections.[18] In particular, the theoretical bounds reveal the dependency of (expected) regret on characteristic parameters of the auction (i.e., T, N, K, and $q_{\min}$, and μ if the mechanism is randomized). Nonetheless, the upper bounds may be inaccurate in overestimating the actual performance of the proposed algorithms. In fact, while we prove that the regret can *never* be larger than the upper bound, some steps in the proofs may be loose, thus leading to bounds which do not accurately predict the behavior of the algorithms in practice. In the following we investigate by means of numerical simulations whether the dependency, in terms of *asymptotic order*, of the bounds on each single parameter of the auction is accurate except for a numerical constant factor. In all the following experiments, we generate the parameters related to the ads in the same way. The qualities $\{q_i\}_{i\in\mathcal{N}}$ are drawn from a uniform distribution in $[0.01, 0.1]$, while the values $\{v_i\}_{i\in\mathcal{N}}$ are randomly drawn from a uniform distribution on $[0, 1]$ ($v_{\max} = 1$). On the other hand, the cumulative probabilities of observation $\{\Lambda_m\}_{m\in\mathcal{K}}$ are different case by case.

Since the main objective is to evaluate the asymptotic accuracy of the bounds, we report the *relative regret*

$$RR_T = \frac{R_T}{B(T, K, N, q_{\min})},$$

where $B(T, K, N, q_{\min})$ is the value of the bound for the specific setting (i.e., Eq. 20 and Eq. 30 for position–dependent, and Eq. 34 for position/ad–dependent externalities).

We analyze the asymptotic accuracy of the bounds w.r.t. each specific parameter, changing only its value and keeping the values of all the others fixed. Since B is proved to be an upper–bound on the actual regret R_T, we expect the relative regret RR_T to be always smaller than 1 ($RR_T = 1$ corresponds to the case in which the experimental regret perfectly matches our upper bound). In particular, we say that our upper–bound accurately predicts the actual asymptotic dependency of the regret w.r.t. a specific parameter if the experimental dependence of RR_T w.r.t. to the parameter is a

[18] Since the bounds over the regret of the SW present a structure similar to those over the auctioneer's revenue, their empirical analysis is omitted.

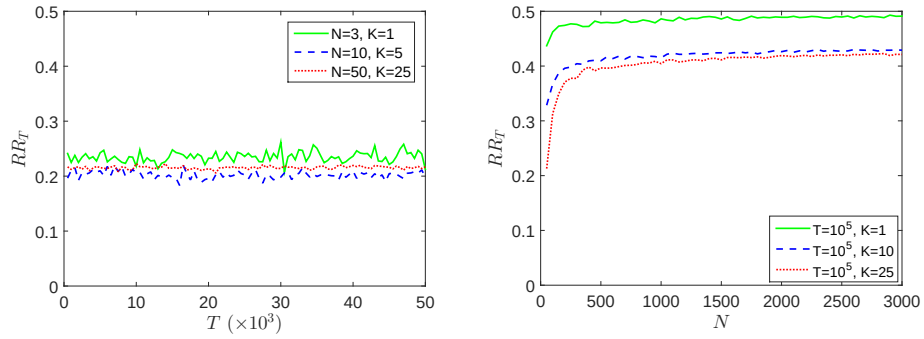

Figure 7: Position–dependent externalities with unknown $\{q_i\}_{i\in\mathcal{N}}$. Dependency of the relative regret on T (left) and N (right).

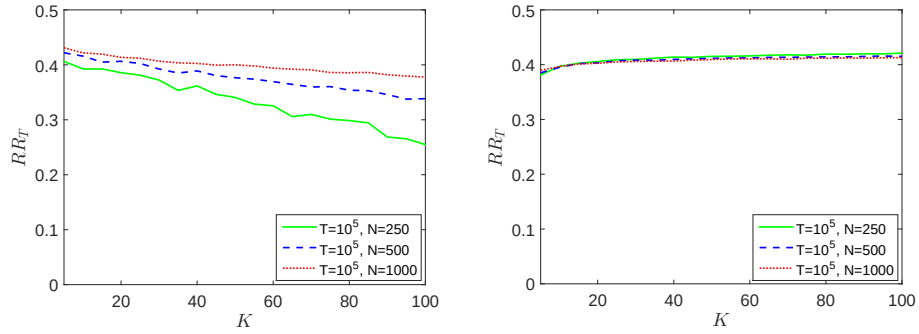

Figure 8: Position–dependent externalities with unknown $\{q_i\}_{i\in\mathcal{N}}$. Dependency of the relative regret on K for two experimental settings (distinguished by the probability distribution according to q are drawn).

constant as the parameter changes. Notice that we do not expect the constant to be close to 1, given that we focus on the asymptotic dependence w.r.t. the parameters and in the steps of the proofs we often use worst–case distribution–free bounds. All the results presented in the following sections are obtained by setting τ and δ as suggested by our bounds and, where it is not differently specified, by averaging over 100 independent runs.

6.1. Position–Dependent Externalities

6.1.1. Unknown $\{q_i\}_{i\in\mathcal{N}}$

First of all we analyze the asymptotic accuracy of the bound provided in Section 4.1, where the model presents only position–dependent externalities and the qualities of the ads are unknown. We design the simulations such

that $\lambda_m = \lambda$ for every m with $\Lambda_1 = 1$ and $\Lambda_K = 0.8$ (i.e., $\lambda = \sqrt[K-1]{0.8}$). Thus, $\Lambda_{\min} = 0.8$ in all the experiments.

In Fig. 7 we analyze the asymptotic accuracy of the bound w.r.t. the parameters T and N. All the three curves in the left plot are completely flat (except for noise due to the randomness of the simulations) showing that the value of the relative regret RR_T for different values of K and N does not change as T increases. This suggests that the bound in Theorem 2 effectively predicts the dependency of the regret R_T w.r.t. the number of steps T of the auction as $\tilde{O}(T^{2/3})$. The right plot represents the dependency of the relative regret RR_T on the number of ads N. In this case we notice that it is relatively accurate as N increases, but there is a transitory effect for smaller values of N where the regret grows faster than predicted by the bound (although $B(T, K, N, q_{\min}, \Lambda_{\min})$ is still an upper–bound to R_T). Finally, the left plot of Fig. 8 suggests that the asymptotic dependency on K in the bound of Theorem 2 is over–estimated, since the relative regret RR_T decreases as K increases. As discussed in the comment to the proof in Section 4, this might be explained by the over–estimation of the term $\frac{\max_i(\tilde{q}_i^+ \hat{v}_i; l)}{\max_i(\tilde{q}_i^+ \hat{v}_i; k)}$ in the proof. In fact, this term is likely to decrease as K increases. In order to validate this intuition, we have identified some experimental settings for which the bound seems to accurately predict the asymptotic dependency on K: $q_1 = 0.1$, $q_2 = 0.095$, and $q_i = 0.09$ for every $2 < i \leq K$. As a result, the ratio between the qualities $\{q_i\}_{i \in \mathcal{N}}$ is fixed (on average) and does not change with K. The right plot of Fig. 8 shows that, with these values of $\{q_i\}_{i \in \mathcal{N}}$, the ratio RR_T is constant for different values of N, implying that in this case the bound accurately predicts the asymptotic behavior of R_T. In fact, as commented in the remarks to Theorem 2, we derive distribution–independent bounds where the qualities $\{q_i\}_{i \in \mathcal{N}}$ do not appear in the bound. As a result, R_T should be intended as a worst case w.r.t. all the possible configurations of qualities and externalities.

6.1.2. Unknown $\{\Lambda_m\}_{m \in \mathcal{K}}$

We now investigate the asymptotic accuracy of the bound derived for algorithm A–VCG2′ presented in Section 4.2.2. We used several probability distributions to generate the values of $\{\lambda_m\}_{m \in \mathcal{K}}$. We observed that, when they are drawn uniformly from the interval $[0.98, 1.00]$, the numerical simulations confirm our bound (as we show below), whereas the bound seems to overestimate the dependencies on K and μ when the support of the proba-

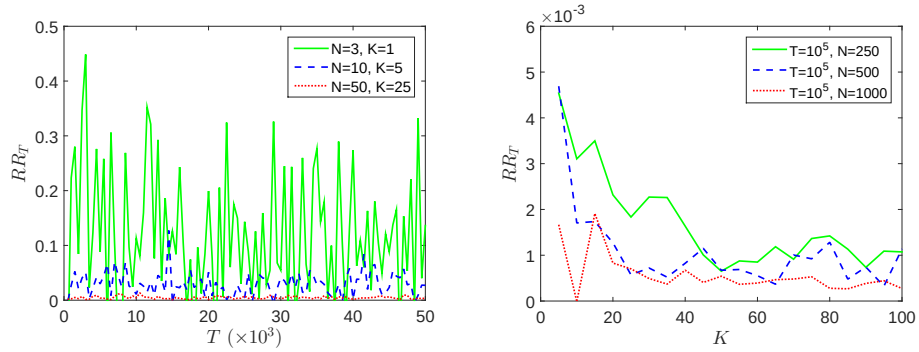

Figure 9: Position–dependent externalities with unknown $\{\Lambda_m\}_{m\in\mathcal{K}}$. Dependency of the relative regret on T (left) and K (right).

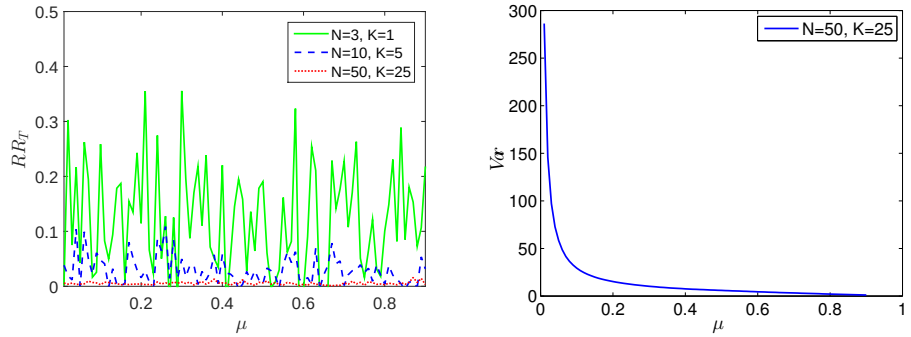

Figure 10: Position–dependent externalities with unknown $\{\Lambda_m\}_{m\in\mathcal{K}}$. Dependency of the relative regret on μ (left). Variance of the revenue of the auctioneer (right).

bility distribution is wider (e.g., $[\xi, 1.00]$ with $\xi \ll 0.98$); we do not report any plot for this second case.

The left plot of Fig. 9 shows the dependence of the ratio RR_T w.r.t. T when $\mu = 0.01$. Despite the noise, the ratio seems not to be affected by the variation of T, confirming our bound. In the right plot of Fig. 9, we observe that when $T = 10^5$ and $\mu = 0.01$ the behavior of the ratio as K changes is essentially the same for different values of N. Furthermore, we observe that the bound is accurate except that it seems to overestimate the dependence when K assumes small values (as it happens in practice). In the left plot of Fig. 10, the ratio RR_T seems to be constant as μ varies when $T = 10^5$.

We conclude our analysis studying the variance of the payments as μ varies. The bound over R_T, provided in Section 4.2.2, suggests to choose

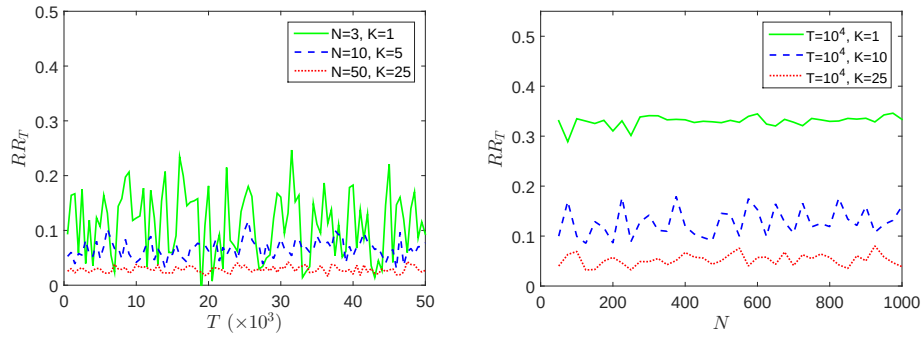

Figure 11: Position–dependent externalities with unknown $\{q_i\}_{i\in\mathcal{N}}$ and $\{\Lambda_m\}_{m\in\mathcal{K}}$. Dependency of the relative regret on T (left) and N (right).

a $\mu \to 0$ in order to reduce the regret. Nonetheless, the regret bounds are obtained in expectation w.r.t. all the sources of randomization and do not consider how single realizations of the learning mechanism may deviate w.r.t. the expected regret. Thus in the right plot of Fig. 10 we investigate the variance of the payments. The variance is excessively high for small values of μ, making the adoption of these value inappropriate. Thus, the choice of μ should consider both these two dimensions of the problem: the regret and the variance of the payments.

6.1.3. Unknown $\{\Lambda_m\}_{m\in\mathcal{K}}$ and $\{q_i\}_{i\in\mathcal{N}}$

In this section we analyze the bound provided in Section 4.3 for position–dependent auctions where both the prominences and the qualities are unknown. For these simulations we generate $\{\lambda_m\}_{m\in\mathcal{K}}$ samples from a uniform distribution over $[0.5, 1]$ and we set τ, δ and μ to the values derived for the bound. In particular, in order to balance the increase of variance of the payments when μ decreases, the number of steps is not constant, but it changes as a function of μ as $\frac{1000}{\mu}$. This means that, in expectation, the bid of a generic ad a_i is modified 1000 times over the number of the steps.

In the plots of Fig. 11, we show that the bound in Eq. 30 accurately predicts the asymptotic dependence of the regret w.r.t. the parameters T and N. Indeed, except for the noise due to the high variance of the payments based on the cSRP, the two plots show that fixing the other parameters, the ratio RR_T is constant as both T increases and N increases.

The plot in Fig. 12 represents the dependency of the relative regret w.r.t. the parameter K. We can deduce that the bound R_T over–estimates the

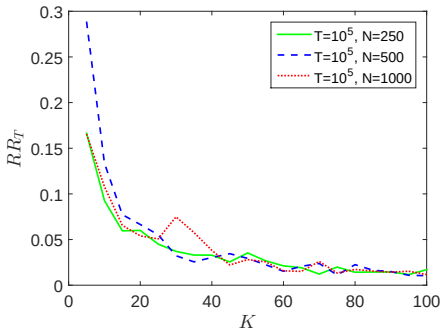

Figure 12: Position–dependent externalities with unknown $\{q_i\}_{i\in\mathcal{N}}$ and $\{\Lambda_m\}_{m\in\mathcal{K}}$. Dependency of the relative regret on K.

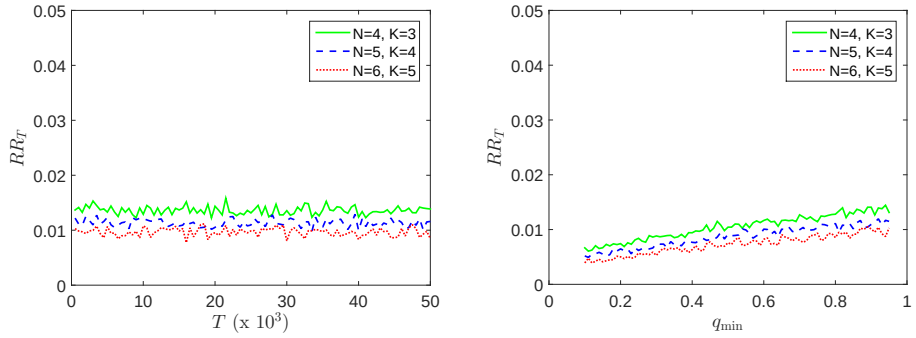

Figure 13: Dependency on T (left) and $q_{\min}$ (right) in auctions with position/ad–dependent externalities.

dependency on K for small values of the parameters, while, with larger values, the bound accurately predicts the behavior, the curves being flat.

6.2. Position/Ad–Dependent Externalities

In this section we analyze the bound provided in Section 5.1 for auctions with position–dependent and ad–dependent externalities where only the qualities are unknown.

In the bound provided in Theorem 14 the regret R_T presents a linear dependency on N and an inverse dependency on the smallest quality $q_{\min}$. In the left plot of Fig. 13 we report RR_T as T increases. As it can be observed, the bound accurately predicts the behavior of the regret w.r.t. T as in the case of position–dependent externalities. In the right plot of Fig. 13 we report RR_T as we change $q_{\min}$. According to the bound in Eq. 34 the regret should decrease as $q_{\min}$ increases (i.e., $R_T = \tilde{O}(q_{\min}^{-1})$) but it is clear from the plot

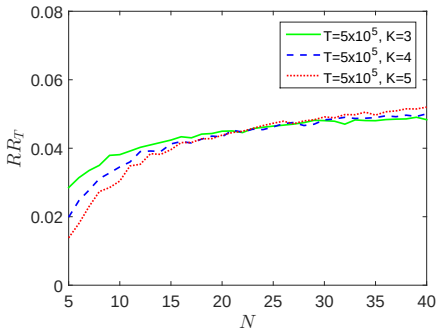

Figure 14: Dependency of the relative regret RR_T on N.

that R_T has a much smaller dependency on $q_{\min}$, if any.[19] Finally, we study the dependency on N (Fig. 14). In this case RR_T slightly increases and then it tends to flat as N increases. This result suggests that the, theoretically derived, $N^{\frac{1}{3}}$ asymptotic dependency of R_T w.r.t. the number of ads might be correct. We do not report results on K since the complexity of finding the optimal allocation f^* becomes intractable for values of K larger than 8, as shown in [41], making the empirical evaluation of the bound unfeasible.

7. Conclusions and Future Work

In this paper, we studied the problem of learning the CTRs of ads in sponsored search auctions with truthful mechanisms. This problem is highly challenging since it requires the combination of online learning tools (i.e., regret minimization algorithms) and economic tools (i.e., truthful mechanisms). While almost all the literature focused on single–slot scenarios, here we focused on multi–slot scenarios. With multiple slots it is necessary to adopt a user model to characterize how the CTR of an ad varies as the allocation of displayed ads varies. Here, we adopted the cascade model, that is the most common model used in the literature. In the paper, we studied a number of scenarios, each with a specific information setting of unknown parameters. For each scenario, we designed a truthful learning mechanism, studied its economic properties, derived an upper bound over the regret, and,

[19] From this experiment is not clear whether $RR_T = \tilde{O}(q_{\min}^{-1})$, thus implying that R_T does not depend on $q_{\min}$ at all, or RR_T is sublinear in $q_{\min}$, which would correspond to a dependency $R_T = \tilde{O}(q_{\min}^{-z})$ with $0 < z < 1$.

for some mechanisms, also a lower bound. We considered both the regret over the auctioneer's revenue and the SW.

We showed that for the cascade model with only position–dependent externalities it is possible to design a truthful no–regret learning mechanism for the general case in which all the parameters are unknown. Our mechanism presents a regret $\tilde{O}(T^{\frac{2}{3}})$ and it is DSIC in expectation w.r.t. the realization of the random component of the mechanism. However, it remains open whether or not it is possible to obtain a regret $\tilde{O}(T^{\frac{1}{2}})$. For specific cases, in which some parameters are known to the auctioneer, we obtained better results in terms of either incentive compatibility, obtaining dominant strategy truthfulness, or regret, obtaining a regret of zero. We showed that for the cascade model with the position– and ad–dependent externalities it is possible to design a DSIC *a posteriori* mechanism with a regret $\tilde{O}(T^{\frac{2}{3}})$ when only the quality is unknown. Instead, even when the cascade model is only with ad–dependent externalities and no parameter is known, it is not possible to obtain a no–regret DSIC *a posteriori* mechanism. The proof of this result would seem to suggest that the same result holds also when a random mechanism is adopted and the truthfulness is in expectation w.r.t. its realizations. However, we did not produce any proof for that, leaving it for future works. Finally, we empirically evaluated the bounds we provided, showing that the dependency of the regret on the parameters is mostly correct in a worst–case scenario.

Two main questions deserve future investigation. The first question concerns the study of a lower bound for the case in which there are only position–dependent externalities and truthfulness is in expectation in expectation w.r.t. only the realizations of the random component of the mechanism or also w.r.t. the click realizations. Furthermore, it is open whether the separation of exploration and exploitation phases is necessary and, in the negative case, whether it is possible to obtain a regret $\tilde{O}(T^{\frac{1}{2}})$. The second question concerns a similar study related to the case with only ad–dependent externalities.

Glossary

AE Allocative Efficiency. 8, 9, 16, 34, 36, 67, 84, 88

BIC Bayesian Incentive Compatibility. 3, 8

cSRP canonical Self–Resampling Procedure. 37, 57, 73, 74, 76, 91

CTR Click–Through–Rate. 1, 2, 3, 5, 11, 15, 17, 18, 20, 23, 25, 40, 59

DSIC Dominant Strategy Incentive Compatibility. 8, 9, 16, 17, 18, 19, 20, 21, 20, 21, 23, 24, 26, 27, 28, 29, 30, 32, 34, 35, 36, 38, 39, 40, 43, 44, 45, 47, 50, 51, 52, 60, 67, 69, 74

EC Execution Contingent. 8

GSP Generalized Second Price. 2, 3, 4

IC Incentive Compatibility. 8, 9, 30

IR Individual Rationality. 8, 9, 16, 17, 18, 21, 26, 28, 33, 34, 36, 38, 43, 47, 67

MAB Multi–Armed Bandit. 3, 4, 7, 10, 11, 17, 24, 29, 30, 32, 41, 45

PoA Price of Anarchy. 3

SSA Sponsored Search Auction. 1, 2, 3, 4, 5, 7, 12, 13, 15, 20, 67

SW Social Welfare. 15, 16, 17, 18, 19, 20, 21, 23, 27, 31, 36, 39, 42, 45, 47, 49, 50, 51, 52, 59, 88, 91

UCB Upper–Confidence Bound. 11, 30

VCG Vickrey–Clarke–Groves. 1, 3, 9, 16, 17, 18, 24, 29, 32, 34, 36, 44, 48, 67, 69, 73, 74, 75, 76, 77, 84, 88

WBB Weak Budget Balance. 8, 9, 16, 18, 21, 28, 34, 35, 36, 38, 43, 48, 67

WVCG Weighted Vickrey–Clarke–Groves. 9, 17, 27, 28, 39, 42, 47, 52, 77

References

[1] IAB, IAB internet advertising revenue report. 2010 first half–year results (2010).

[2] H. R. Varian, C. Harris, The VCG auction in theory and practice, American Economic Review 104 (5) (2014) 442–445.

[3] Y. Narahari, D. Garg, R. Narayanam, H. Prakash, Game Theoretic Problems in Network Economics and Mechanism Design Solutions, Springer, 2009.

[4] B. Edelman, M. Ostrovsky, M. Schwarz, Internet advertising and the generalized second–price auction: Selling billions of dollars worth of keywords, American Economic Review 97 (1) (2007) 242–259.

[5] H. R. Varian, Position auctions, International Journal of Industrial Organization 25 (6) (2007) 1163–1178.

[6] J. Hegeman, Facebook's ad auction, Talk at Ad Auctions Workshop.

[7] R. Leme, E. Tardos, Pure and bayes-nash price of anarchy for generalized second price auction, in: Proceedings of FOCS, 2010, pp. 735–744.

[8] E. T. Brendan Lucier, Renato Paes Leme, On revenue in the generalized second price auction, in: Proceedings of WWW, 2012, pp. 361–370.

[9] E. Markakis, O. Telelis, Discrete strategies in keyword auctions and their inefficiency for locally aware bidders, in: Proceedings of the International Workshop on Internet and Network Economics (WINE), 2010, pp. 523–530.

[10] D. Kuminov, M. Tennenholtz, User modeling in position auctions: reconsidering the gsp and vcg mechanisms, in: Proceedings of AAMAS, 2009, pp. 273–280.

[11] R. Gomes, N. Immorlica, E. Markakis, Externalities in keyword auctions: An empirical and theoretical assessment, in: Proceedings of the International Workshop on Internet and Network Economics (WINE), 2009, pp. 172–183.

[12] B. Yoram, S. Ceppi, I. A. Kash, P. Key, D. Kurokawa, Optimising trade-offs among stakeholders in ad auctions, in: Proceedings of the International ACM Conference on Economics and Computation (EC), 2014, pp. 75–92.

[13] L. Tran-Thanh, S. Stein, A. Rogers, N. R. Jennings, Efficient crowdsourcing of unknown experts using multi–armed bandits, Artificial Intelligence Journal 214 (2014) 89 –111.

[14] H. Robbins, Some aspects of the sequential design of experiments, Bulletin of the AMS 58 (1952) 527–535.

[15] S. Pandey, C. Olston, Handling Advertisements of Unknown Quality in Search Advertising, in: Proceedings of the Conference on Neural Information Processing Systems (NIPS), 2006, pp. 1065–1072.

[16] J. Langford, L. Li, Y. Vorobeychik, J. Wortman, Maintaining equilibria during exploration in sponsored search auctions, Algorithmica 58 (2010) 990–1021.

[17] R. Gonen, E. Pavlov, An incentive-compatible multi–armed bandit mechanism, in: ACM Symp. on Principles Of Distributed Computing (PODC) (Brief Announcement), pages 362–363, 2007. Preliminary version in 3rd Workshop on Sponsored Search Auctions (in conjunction with WWW 2007).

[18] R. Gonen, E. Pavlov, Adaptive incentive–compatible sponsored search auction, in: Conference on Current Trends in Theory and Practice of Computer Science, 2009, pp. 303–316.

[19] H. Nazerzadeh, A. Saberi, R. Vohra, Dynamic cost-per-action mechanisms and applications to online advertising, in: Proceeding of the International Conference on World Wide Web (WWW), 2008, pp. 179–188.

[20] N. R. Devanur, S. M. Kakade, The price of truthfulness for pay-per-click auctions, in: Proceedings of the ACM Conference on Electronic Commerce (ACM EC), 2009, pp. 99–106.

[21] M. Babaioff, Y. Sharma, A. Slivkins, Characterizing truthful multi–armed bandit mechanisms: Extended abstract, in: Proceedings of the

ACM Conference on Electronic Commerce (ACM EC), ACM, 2009, pp. 79–88.

[22] R. P. M. Sai-Ming Li, Mohammad Mahdian, Value of learning in sponsored search auctions, in: Proceedings of the International Workshop on Internet and Network Economics (WINE), 2010, pp. 294–305.

[23] M. Babaioff, R. D. Kleinberg, A. Slivkins, Truthful mechanisms with implicit payment computation, in: Proceedings of the ACM Conference on Electronic Commerce (EC), 2010, pp. 43–52.

[24] D. Kempe, M. Mahdian, A cascade model for externalities in sponsored search, in: Proceedings of the International Workshop on Internet and Network Economics (WINE), 2008, pp. 585–596.

[25] G. Aggarwal, J. Feldman, S. Muthukrishnan, M. Pál, Sponsored search auctions with markovian users, in: Proceedings of the International Workshop on Internet and Network Economics (WINE), 2008, pp. 621–628.

[26] N. Craswell, O. Zoeter, M. Taylor, B. Ramsey, An experimental comparison of click position–bias models, in: Proceedings of the International Conference Web Search and Web Data Mining (WSDM), 2008, pp. 87–94.

[27] T. Joachims, L. Granka, B. Pan, H. Hembrooke, F. Radlinski, G. Gay, Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search, ACM Transactions on Information Systems 25 (2) (2007) 7.

[28] A. D. Sarma, S. Gujar, Y. Narahari, Truthful multi–armed bandit mechanisms for multi–slot sponsored search auctions, Current Science, Special Issue on Game Theory 103 (9) (2012) 1064–1077.

[29] N. Gatti, A. Lazaric, F. Trovò, A truthful learning mechanism for contextual multi–slot sponsored search auctions with externalities, in: Proceedings of the ACM Conference on Electronic Commerce (ACM EC), 2012, pp. 605–622.

[30] A. Mas-Colell, M. Whinston, J. Green, Microeconomic theory, Oxford university press New York, 1995.

[31] E. H. Gerding, S. Stein, K. Larson, A. Rogers, N. R. Jennings, Scalable mechanism design for the procurement of services with uncertain durations, in: Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2010, pp. 649–656.

[32] S. Ceppi, N. Gatti, E. H. Gerding, Mechanism design for federated sponsored search auctions, in: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2011, pp. 608–613.

[33] N. Nisan, A. Ronen, Computationally feasible vcg mechanisms, Journal of Artificial Intelligence Research 29 (1) (2007) 19–47.

[34] N. Nisan, T. Roughgarden, E. Tardos, V. V. Vazirani, Algorithmic Game Theory, Cambridge University Press, 2007.

[35] A. Archer, E. Tardos, Truthful mechanisms for one–parameter agents, in: Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS), 2001, pp. 482–491.

[36] R. B. Myerson, Optimal auction design, Mathematics of Operations Research 21 (6) (1981) 58–73.

[37] A. Archer, C. Papadimitriou, K. Talwar, E. Tardos, An approximate truthful mechanism for combinatorial auctions with single parameter agents, in: Proceedings of the ACM–SIAM Symposium on Discrete Algorithms (SODA), 2003, pp. 205–214.

[38] J. Gittins, Bandit processes and dynamic allocation indices, Journal of the Royal Statistical Society **41** (1979) 148–164.

[39] S. Bubeck, N. Cesa-Bianchi, Regret analysis of stochastic and nonstochastic multi–armed bandit problems, Foundations and Trends in Machine Learning 5 (1) (2012) 1–122.

[40] P. Auer, N. Cesa-Bianchi, P. Fischer, Finite-time analysis of the multiarmed bandit problem, Machine Learning 47 (2-3) (2002) 235–256.

[41] N. Gatti, M. Rocco, Which mechanism in sponsored search auctions with externalities?, in: Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2013, pp. 635–642.

[42] P. Auer, N. Cesa-Bianchi, P. Fischer, Finite–time analysis of the multi-armed bandit problem, Machine Learning Journal 47 (2002) 235–256.

[43] D. Mandal, Y. Narahari, A novel ex-post truthful mechanism for multi-slot sponsored search auctions, in: International conference on Autonomous Agents and Multi-Agent Systems, AAMAS '14, Paris, France, May 5-9, 2014, 2014, pp. 1555–1556.
URL http://dl.acm.org/citation.cfm?id=2616059

[44] W. Hoeffding, Probability inequalities for sums of bounded random variables, Journal of the American Statistical Association 58 (1963) 13–30.

[45] J. R. Green, J.-J. Laffont, Incentives in Public Decision Making, North-Holland, 1979.

Appendix A. Vickrey–Clarke–Groves mechanism

Consider a generic direct–revelation mechanism $M = (\mathcal{N}, \mathcal{V}, \Theta, f, \{p_i\}_{i\in\mathcal{N}})$ as defined in Section 3.2. Differently from the SSA case, in general the type of an agent, denoted by v_i for consistency with the rest of the paper, is a vector of parameters. We define a function $val_i : \Theta \times \mathcal{V} \to \mathbb{R}^+$, which returns the value obtained by agent a_i when its type is v_i and the allocation chosen by the mechanism is θ.

The VCG mechanism is obtained coupling the two following functions:

- the allocation function f which returns the allocation maximising the social welfare, i.e.,

$$f(\hat{\mathbf{v}}) = \arg\max_{\theta\in\Theta} SW(\theta, \hat{\mathbf{v}}) = \arg\max_{\theta\in\Theta} \sum_{i\in\mathcal{N}} val_i(\theta, \hat{v}_i);$$

- the payment rule p_i, which defines the payment required from agent a_i, i.e.,

$$\begin{aligned} p_i(\hat{\mathbf{v}}) &= SW(f(\hat{\mathbf{v}}_{-i}), \hat{\mathbf{v}}_{-i}) - SW_{-i}(f(\hat{\mathbf{v}}), \hat{\mathbf{v}}) \\ &= \sum_{j\in\mathcal{N}, j\neq i} val_j(f(\hat{\mathbf{v}}_{-i}), \hat{v}_j) - \sum_{j\in\mathcal{N}, j\neq i} val_j(f(\hat{\mathbf{v}}), \hat{v}_j), \end{aligned}$$

  where we denote by $f(\hat{\mathbf{v}}_{-i})$ the allocation returned by f when agent i does not participate to the auction.

In this quasi–linear environment, when there are no interdependencies among the types of the agents and the no–single–agent effect [3] holds, the VCG mechanism is AE, DSIC *a posteriori*, IR *a posteriori*, and WBB *a posteriori*.

Appendix B. Monotonicity and Myerson's payments

Consider a generic direct–revelation mechanism $M = (\mathcal{N}, \mathcal{V}, \Theta, f, \{p_i\}_{i\in\mathcal{N}})$ as defined in Section 3.2. A single–parameter linear environment is such that

- the type of each agent a_i is a scalar v_i (single–parameter assumption),
- the utility function of agent a_i is $u_i(\hat{\mathbf{v}}) = z_i\left(f\left(\hat{\mathbf{v}}\right)\right) v_i - p_i(\hat{\mathbf{v}})$ where $z_i : \Theta \to \mathbb{R}$ is a function of the allocation (linear assumption).

An allocation function f is *monotonic* in a single–parameter linear environment if for any $\hat{\mathbf{v}}_{-i}$

$$z_i\left(f\left(\hat{\mathbf{v}}_{-i}, v_i''\right)\right) \geq z_i\left(f\left(\hat{\mathbf{v}}_{-i}, v_i'\right)\right) \quad \forall i \in \mathcal{N}$$

for any $v_i'' \geq v_i'$. Essentially, z_i is monotonically increasing in v_i once $\hat{\mathbf{v}}_{-i}$ has been fixed. In such environments, it is always possible to design a DSIC mechanism imposing the following payments [35]:

$$p_i(\hat{\mathbf{v}}) = h_i(\hat{\mathbf{v}}_{-i}) + z_i\left(f\left(\hat{\mathbf{v}}\right)\right)\hat{v}_i - \int_0^{\hat{v}_i} z_i\left(f\left(\hat{\mathbf{v}}_{-i}, u\right)\right) du \tag{B.1}$$

where $h_i : \mathcal{V}^{N-1} \rightarrow \mathbb{R}$ is a generic function not depending on the type of agent a_i.

Appendix C. Proof of Revenue Regret in Theorem 2

We start by reporting the proof of Proposition 1.

Proof. *(Proposition 1)* The derivation is a simple application of the Hoeffding's bound. We first notice that each of the terms in the empirical average $\tilde{q}_i$ (Eq. 11) is bounded in $[0; 1/\Lambda_{\pi(i;\theta_t)}]$. Thus we obtain

$$\mathbb{P}\left(|q_i - \tilde{q}_i| \geq \epsilon\right) \leq 2\exp\left(-\frac{2|B_i|^2\epsilon^2}{\sum_{t\in B_i}\left(\frac{1}{\Lambda_{\pi(i;\theta_t)}} - 0\right)^2}\right) = \frac{\delta}{N}.$$

By reordering the terms in the previous expression we have

$$\eta = \sqrt{\left(\sum_{t\in B_i}\frac{1}{\Lambda^2_{\pi(i;\theta_t)}}\right)\frac{1}{2|B_i|^2}\log\frac{2N}{\delta}},$$

which guarantees that all the empirical estimates $\tilde{q}_i$ are within η of q_i for all the ads with probability, at least, $1-\delta$. □

Before stating the main result of this section, we need the following lemma.

Lemma 1. *For any slot s_m with $m \in \mathcal{K}$, with probability $1-\delta$,*

$$\frac{\max\limits_{i\in\mathcal{N}}(q_i\hat{v}_i;m)}{\max\limits_{i\in\mathcal{N}}(\tilde{q}_i^{+}\hat{v}_i;m)} \leq 1, \tag{C.1}$$

where the operator $\max(\cdot;\cdot)$ *is defined as in Section 4.*

Proof. The proof is a straightforward application of Proposition 1. We consider the optimal allocation θ^* defined in Eq. 2 and the optimal allocation $\tilde{\theta}$ when estimates $\tilde{q}^{+}$ are adopted defined in Eq. 16. We denote $h = \alpha(m;\theta^*) \in \arg\max\limits_{i\in\mathcal{N}}(q_i\hat{v}_i;m)$, i.e., the index of the ad allocated in a generic slot in position m. There are two possible scenarios:

- If $\pi(h;\tilde{\theta}) < m$ (the ad is displayed into a higher slot in the approximated allocation $\tilde{\theta}$), then $\exists j \in \mathcal{N}$ s.t. $\pi(j;\theta^*) < m \wedge \pi(j;\tilde{\theta}) \geq m$. Thus

$$\max_{i\in\mathcal{N}}(\tilde{q}_i^{+}\hat{v}_i;m) \geq \tilde{q}_j^{+}\hat{v}_j \geq q_j\hat{v}_j \geq q_h\hat{v}_h = \max_{i\in\mathcal{N}}(q_i\hat{v}_i;m)$$

  where the second inequality holds with probability $1-\delta$;

- If $\pi(h;\tilde{\theta}) \geq m$ (the ad is displayed into a lower or equal slot in the approximated allocation $\tilde{\theta}$), then

$$\max_{i\in\mathcal{N}}(\tilde{q}_i^{+}\hat{v}_i;m) \geq \tilde{q}_h^{+}\hat{v}_h \geq q_h v_h = \max_{i\in\mathcal{N}}(q_i\hat{v}_i;m)$$

  where the second inequality holds with probability $1-\delta$.

In both cases, the statement follows. □

Proof. *(Theorem 2)*
Step 1: expected payments. The proof follows steps similar to those in the proofs in [20]. We first recall that since the mechanism is DSIC in expectation w.r.t. the clicks, then we can directly focus on the regret when the actual values $\mathbf{v}$ are bid. For any ad a_i such that $\pi(i;\theta^*) \leq K$, the expected payments of the VCG mechanism in this case reduce to Eq. 9:

$$p_i^*(\mathbf{v}) = \sum_{l=\pi(i;\theta^*)+1}^{K+1} \left[(\Lambda_{l-1} - \Lambda_l) \max_{j\in\mathcal{N}}(q_j v_j; l) \right],$$

while, given the definition of A–VCG1 reported in Section 4.1, the expected payments for at t–th iteration of the auction are

$$\tilde{p}_{i,t}(\mathbf{v}) = \begin{cases} 0 & \text{if } t \leq \tau \ (\textit{exploration}) \\ \tilde{p}_i(\mathbf{v}) & \text{if } t > \tau \ (\textit{exploitation}) \end{cases} \tag{C.2}$$

where the payment for any ad a_i such that $\pi(i;\tilde{\theta}) \leq K$ is defined in Eq. 17 as:

$$\tilde{p}_i(\mathbf{v}) = \frac{q_i}{\tilde{q}_i^+} \sum_{l=\pi(i;\tilde{\theta})+1}^{K+1} (\Lambda_{l-1} - \Lambda_l) \max_{j \in \mathcal{N}}(\tilde{q}_j^+ v_j; l).$$

Step 2: per–step exploration regret. Since for any $1 \leq t \leq \tau$, A–VCG1 sets all the payments to 0, the per–step regret is

$$r_t = \sum_{m=1}^{K} \left(p^*_{\alpha(m;\theta^*)}(\mathbf{v}) - 0 \right) = \sum_{m=1}^{K} \sum_{l=m}^{K} \Delta_l \max_{i \in \mathcal{N}}(q_i v_i; l+1) \leq v_{\max} \sum_{m=1}^{K} \Lambda_m, \tag{C.3}$$

where $\Delta_l = \Lambda_l - \Lambda_{l+1}$. The exploration regret is obtained by summing up r over τ steps.

Step 3: per–step exploitation regret. Now we focus on the expected (w.r.t. click realizations) per–step regret during the exploitation phase. According to the definition of payments, at each step $t \in \{\tau+1, \ldots, T\}$ of the exploitation phase we bound the per–step regret r as

$$\begin{aligned}
r_t &= \sum_{m=1}^{K} \left(p^*_{\alpha(m;\theta^*)}(\mathbf{v}) - \tilde{p}_{\alpha(m;\tilde{\theta})}(\mathbf{v}) \right) \\
&= \sum_{m=1}^{K} \sum_{l=m}^{K} \Delta_l \left(\max_{i \in \mathcal{N}}(q_i v_i; l+1) - \frac{\max_{i \in \mathcal{N}}(\tilde{q}_i^+ v_i; l+1)}{\tilde{q}^+_{\alpha(m;\tilde{\theta})}} q_{\alpha(m;\tilde{\theta})} \right) \\
&= \sum_{m=1}^{K} \sum_{l=m}^{K} \Delta_l \frac{\max_{i \in \mathcal{N}}(\tilde{q}_i^+ v_i; l+1)}{\tilde{q}^+_{\alpha(m;\tilde{\theta})}} \left(\frac{\max_{i \in \mathcal{N}}(q_i v_i; l+1)}{\max_{i \in \mathcal{N}}(\tilde{q}_i^+ v_i; l+1)} \tilde{q}^+_{\alpha(m;\tilde{\theta})} - q_{\alpha(m;\tilde{\theta})} \right) \\
&= \sum_{m=1}^{K} \sum_{l=m}^{K} \Delta_l \frac{\max_{i \in \mathcal{N}}(\tilde{q}_i^+ v_i; l+1)}{\max_{i \in \mathcal{N}}(\tilde{q}_i^+ v_i; m)} v_{\alpha(m;\tilde{\theta})} \left(\frac{\max_{i \in \mathcal{N}}(q_i v_i; l+1)}{\max_{i \in \mathcal{N}}(\tilde{q}_i^+ v_i; l+1)} \tilde{q}^+_{\alpha(m;\tilde{\theta})} - q_{\alpha(m;\tilde{\theta})} \right).
\end{aligned}$$

By definition of the max operator, since $l+1 > m$, it follows that

$$\frac{\max_{i \in \mathcal{N}}(\tilde{q}_i^+ v_i; l+1)}{\max_{i \in \mathcal{N}}(\tilde{q}_i^+ v_i; m)} \leq 1. \tag{C.4}$$

Finally, from Lemma 1 and $v_{\alpha(m;\tilde{\theta})} \leq v_{\max}$, it follows that

$$r_t \leq \sum_{m=1}^{K} \sum_{l=m}^{K} v_{\max} \Delta_l \left(\tilde{q}^+_{\alpha(m;\tilde{\theta})} - q_{\alpha(m;\tilde{\theta})} \right) \leq v_{\max} \sum_{m=1}^{K} \left[\left(\tilde{q}^+_{\alpha(m;\tilde{\theta})} - q_{\alpha(m;\tilde{\theta})} \right) \sum_{l=m}^{K} \Delta_l \right], \tag{C.5}$$

with probability at least $1-\delta$. Notice that, by definition of Δ_l, $\sum_{l=m}^{K} \Delta_l = \Lambda_m - \Lambda_{K+1} = \Lambda_m$. Furthermore, from the definition of $\tilde{q}_i^+$ and using Eq. 14 we have that for any ad a_i:

$$\tilde{q}_i^+ - q_i = \tilde{q}_i - q_i + \eta \leq 2\eta,$$

with probability at least $1-\delta$. Thus, the difference between the payments becomes

$$r_t \leq 2 v_{\max} \left(\sum_{m=1}^{K} \Lambda_m \right) \eta = 2 v_{\max} \left(\sum_{m=1}^{K} \Lambda_m \right) \sqrt{\left(\sum_{m=1}^{K} \frac{1}{\Lambda_m^2} \right) \frac{N}{K^2 \tau} \log \frac{N}{\delta}} \tag{C.6}$$

with probability $1-\delta$.[20]

Step 4: cumulative regret. We first consider the (low-probability) event in which the bound on $\tilde{q}_i^+$ derived in Proposition 1. In this case, we cannot guarantee anything about the behavior of the mechanism, since the payments are very inaccurate estimates of the CTRs, and thus the largest possible regret is suffered. In particular, we consider the worst case loss of $v_{\max}$ for each slot for each step, leading to a total regret of $v_{\max} \left(\sum_{m=1}^{K} \Lambda_m \right) T$ with probability δ. By summing up the regrets reported in Eq. C.3 during the exploration phase and Eq. C.6 during the exploitation phase and by

[20] Notice that in the logarithmic term the factor of 2 we have in Proposition 1 disappears since in this proof we only need the one–sided version of the bound.

considering that these bounds hold with probability at least $1-\delta$ (upper-bounded by 1 in the following), we obtain an expected regret

$$R_T \leq v_{\max}\left(\sum_{m=1}^{K}\Lambda_m\right)\Bigg(\underbrace{2(T-\tau)\sqrt{\left(\sum_{m=1}^{K}\frac{1}{\Lambda_m^2}\right)\frac{N}{K^2\tau}\log\frac{N}{\delta}}}_{R_{ei}}+\underbrace{\tau}_{R_{er}}+\underbrace{\delta T}_{R_\delta}\Bigg),$$

where R_{ei} is the upper bound on the regret suffered during the exploitation phase (which holds with probability at least $1-\delta$), R_{er} is the upper bound on the regret suffered during the exploitation phase (which holds with probability at least $1-\delta$) and R_δ is the upper bound on the regret when the bounds do not hold (with probability at most δ). This boundcan be further simplified, given that $\sum_{m=1}^{K}\Lambda_m \leq K$, as

$$R_T \leq v_{\max}K\left(2(T-\tau)\sqrt{\left(\sum_{m=1}^{K}\frac{1}{\Lambda_m^2}\right)\frac{N}{K^2\tau}\log\frac{N}{\delta}}+\tau+\delta T\right). \tag{C.7}$$

Step 5: parameters optimization. Beside describing the performance of A–VCG1, the previous bound also provides guidance for the optimization of the parameters τ and δ. We first simplify the bound in Eq. C.7 as

$$\begin{aligned} R_T &\leq v_{\max}K\left(2T\sqrt{\left(\sum_{m=1}^{K}\frac{1}{\Lambda_m^2}\right)\frac{N}{K^2\tau}\log\frac{N}{\delta}}+\tau+\delta T\right) \\ &\leq v_{\max}K\left(\frac{2T}{\Lambda_{\min}}\sqrt{\frac{N}{K\tau}\log\frac{N}{\delta}}+\tau+\delta T\right), \end{aligned} \tag{C.8}$$

where we used $\sum\limits_{m=1}^{K}1/\Lambda_m^2 \leq K/\Lambda_{\min}^2$, with $\Lambda_{\min}=\min_{m\in\mathcal{K}}\Lambda_m$. In order to find the optimal value of τ, we take the derivative of the previous bound w.r.t. τ, set it to zero and obtain

$$v_{\max}K\left(-\tau^{-\frac{3}{2}}\frac{T}{\Lambda_{\min}}\sqrt{\frac{N}{K}\log\frac{N}{\delta}}+1\right)=0,$$

which leads to

$$\tau = K^{-\frac{1}{3}}T^{\frac{2}{3}}N^{\frac{1}{3}}\Lambda_{\min}^{-\frac{2}{3}}\Big(\log\frac{N}{\delta}\Big)^{\frac{1}{3}}.$$

Substituting this value of τ into Eq. C.8 leads to the optimized bound

$$R_T \leq v_{\max} K \left(3K^{-\frac{1}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \Lambda_{\min}^{-\frac{2}{3}} \Big(\log \frac{N}{\delta} \Big)^{\frac{1}{3}} + \delta T \right).$$

We are now left with the choice of the confidence parameter $\delta \in (0,1)$, which can be easily set to optimize the asymptotic rate (i.e., ignoring constants and logarithmic factors) as

$$\delta = K^{-\frac{1}{3}} T^{-\frac{1}{3}} N^{\frac{1}{3}}$$

We thus obtain the final bound

$$R_T \leq 4 v_{\max} \Lambda_{\min}^{-\frac{2}{3}} K^{\frac{2}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \left[\log \left(K^{\frac{1}{3}} T^{\frac{1}{3}} N^{\frac{2}{3}} \right) \right]^{\frac{1}{3}}.$$

We have to impose the constraints that $T > \frac{N}{K}$ (given by $\delta < 1$) and that $T > \tau$, i.e., $T > \frac{N}{K\Lambda_{\min}^2} \log \frac{N}{\delta}$. The two constraints imply:

$$T > \frac{N}{K\Lambda_{\min}^2} \max \left\{ \log \frac{N}{\delta}, 1 \right\}.$$

□

Appendix D. Proof of Revenue Regret in Theorem 7

Unlike the setting considered in Theorem 2, here the regret is only due to the use of a randomized mechanism, since no parameter estimation is actually needed.

Proof. *(Theorem 7)*
Step 1: payments and additional notation. We recall that according to [35] and [45] the expected VCG payments can be written, as in Eq. 24, in the form

$$p_i^*(\hat{\mathbf{v}}) = \Lambda_{\pi(i;f^*(\hat{\mathbf{v}}))} q_i \hat{v}_i - \int_0^{\hat{v}_i} \Lambda_{\pi(i;f^*(\hat{\mathbf{v}}_{-i},u))} q_i du,$$

while the A–VCG2′ mechanism prescribes contingent payments as in Eq. 25, which lead to expected payments

$$p_i^{*\prime}(\hat{\mathbf{v}}) = \mathbb{E}_{\mathbf{x}} \big[\Lambda_{\pi(i;f^*(\mathbf{x}))} | \hat{\mathbf{v}} \big] q_i \hat{v}_i - \int_0^{\hat{v}_i} \mathbb{E}_{\mathbf{x}} \big[\Lambda_{\pi(i;f^*(\mathbf{x}))} | \hat{\mathbf{v}}_{-i}, u \big] q_i du. \tag{D.1}$$

Given the randomness of the allocation function of A–VCG2′, we need to introduce the following additional notation:

- $\mathbf{s} \in \{0,1\}^N$ is a vector where each element s_i denotes whether the i–th bid has been preserved or it has been modified by the cSRP, i.e., if $x_i = \hat{v}_i$ then $s_i = 1$, otherwise if $x_i < \hat{v}_i$ then $s_i = 0$. Notice that $\mathbf{s}$ does not provide information about the actual modified values $\mathbf{x}$;

- $\mathbb{E}_{\mathbf{x}|\mathbf{s}}[\Lambda_{\pi(i;f(\mathbf{x}))}|\hat{\mathbf{v}}]$ is the expected value of prominence associated with the slots allocated to ad a_i, conditioned on the declared bids $\hat{\mathbf{v}}$ being perturbed as in $\mathbf{s}$.

Moreover, let $S = \{\mathbf{s}|\pi(i; f^*(\hat{\mathbf{v}})) \leq K+1 \Rightarrow s_i = 1 \ \forall i \in \mathcal{N}\}$ be all the realizations where the cSRP does not modify the bids of the first $K+1$ ads, i.e., the K ads displayed applying f^* to the true bids $\hat{\mathbf{v}}$ and the first non-allocated ad.

Step 2: cumulative regret. We proceed by studying the per–ad regret $r_i(\mathbf{v}) = p_i^*(\mathbf{v}) - p_i^{*\prime}(\mathbf{v})$, where the advertisers bid their true values $\mathbf{v}$ since the mechanism is DSIC. Given the previous definitions, we rewrite the expected payments $p_i^{*\prime}(\mathbf{v})$ as

$$
\begin{aligned}
p_i^{*\prime}(\mathbf{v}) &= \Bigg(\mathbb{P}[\mathbf{s} \in S]\Lambda_{\pi(i;f^*(\mathbf{v}))} + \mathbb{P}[\mathbf{s} \notin S]\mathbb{E}_{\mathbf{x}|\mathbf{s}\notin S}[\Lambda_{\pi(i;f^*(\mathbf{x}))}|\mathbf{v}]\Bigg) q_i v_i + \\
&\quad - \int_0^{v_i} \Bigg(\mathbb{P}[\mathbf{s} \in S]\Lambda_{\pi(i;f^*(\mathbf{v}_{-i},u))} + \mathbb{P}[\mathbf{s} \notin S]\mathbb{E}_{\mathbf{x}|\mathbf{s}\neq\mathbf{1}}[\Lambda_{\pi(i;f^*(\mathbf{x}))}|\mathbf{v}_{-i},u]\Bigg) q_i du \\
&= \mathbb{P}[\mathbf{s} \in S]\Bigg(\Lambda_{\pi(i;f^*(\mathbf{v}))} q_i v_i - \int_0^{v_i} \Lambda_{\pi(i;f^*(\mathbf{v}_{-i},u))} q_i du\Bigg) + \\
&\quad + \mathbb{P}[\mathbf{s} \notin S]\Bigg(\mathbb{E}_{\mathbf{x}|\mathbf{s}\notin S}[\Lambda_{\pi(i;f^*(\mathbf{x}))}|\mathbf{v}] q_i v_i - \int_0^{v_i} \mathbb{E}_{\mathbf{x}|\mathbf{s}\notin S}[\Lambda_{\pi(i;f^*(\mathbf{x}))}|\mathbf{v}_{-i},u] q_i du\Bigg) \\
&= \mathbb{P}[\mathbf{s} \in S] p_i^*(\mathbf{v}) + \\
&\quad + \mathbb{P}[\mathbf{s} \notin S]\Bigg(\mathbb{E}_{\mathbf{x}|\mathbf{s}\notin S}[\Lambda_{\pi(i;f^*(\mathbf{x}))}|\mathbf{v}] q_i v_i - \int_0^{v_i} \mathbb{E}_{\mathbf{x}|\mathbf{s}\notin S}[\Lambda_{\pi(i;f^*(\mathbf{x}))}|\mathbf{v}_{-i},u] q_i du\Bigg),
\end{aligned}
$$

where in the last expression we used the expression of the VCG payments in Eq. 24 according to [35] and [45]. The per–ad regret is

$$
\begin{aligned}
r_i(\mathbf{v}) &= p_i^*(\mathbf{v}) - p_i^{*\prime}(\mathbf{v}) \\
&= p_i^*(\mathbf{v}) - \mathbb{P}[\mathbf{s} \in S] p_i^*(\mathbf{v}) + \\
&\quad - \mathbb{P}[\mathbf{s} \notin S]\Bigg(\mathbb{E}_{\mathbf{x}|\mathbf{s}\notin S}[\Lambda_{\pi(i;f^*(\mathbf{x}))}|\mathbf{v}] q_i v_i - \int_0^{v_i} \mathbb{E}_{\mathbf{x}|\mathbf{s}\notin S}[\Lambda_{\pi(i;f^*(\mathbf{x}))}|\mathbf{v}_{-i},u] q_i du\Bigg)
\end{aligned}
$$

$$= \mathbb{P}[\mathbf{s} \notin S] p_i^*(\mathbf{v}) +$$
$$- \mathbb{P}[\mathbf{s} \notin S] \underbrace{\left(\mathbb{E}_{\mathbf{x}|\mathbf{s}\notin S}[\Lambda_{\pi(i;f^*(\mathbf{x}))}|\mathbf{v}] q_i v_i - \int_0^{v_i} \mathbb{E}_{\mathbf{x}|\mathbf{s}\notin S}[\Lambda_{\pi(i;f^*(\mathbf{x}))}|\mathbf{v}_{-i}, u] q_i du \right)}_{r_{i,1}^B}.$$

Since we have that $u \leq v_i$ in the integral and since the allocation function defined in [23] is monotonic, we have that

$$\mathbb{E}_{\mathbf{x}|\mathbf{s}\notin S}[\Lambda_{\pi(i;f^*(\mathbf{x}))}|\mathbf{v}_{-i}, u] \leq \mathbb{E}_{\mathbf{x}|\mathbf{s}\notin S}[\Lambda_{\pi(i;f^*(\mathbf{x}))}|\mathbf{v}],$$

which implies that $r_{i,1}^B$ is non–negative. Thus the regret r_i^B can be bounded as

$$\begin{aligned} r_i^B(\mathbf{v}) &= \mathbb{P}[\mathbf{s} \notin S] p_i^*(\mathbf{v}) \underbrace{- \mathbb{P}[\mathbf{s} \notin S] r_{i,1}^B}_{\leq 0} \\ &\leq \mathbb{P}[\mathbf{s} \notin S] p_i^*(\mathbf{v}) \leq \mathbb{P}\big[\exists j : s_j = 0 \wedge \pi(j; f^*(\mathbf{v})) \leq K+1\big] v_{\max} \\ &\leq \sum_{j \in \mathcal{N}: \pi(j;f^*(\mathbf{v})) \leq K+1} \mathbb{P}[s_j = 0] v_{\max} \\ &= (K+1)\, \mu v_{\max} \leq 2K \mu v_{\max}. \end{aligned} \tag{D.2}$$

We can now compute the bound on the global regret R_T. Since this mechanism does not require any estimation phase, the regret is simply

$$R_T \leq 2K^2 \mu v_{\max} T.$$

Step 3: parameters optimization. In this case, the bound would suggest to choose a $\mu \to 0$, but it is necessary to consider that with $\mu \to 0$ the variance of the payment goes to infinity.

Appendix E. Proof of Revenue Regret in Theorem 11

The proof of Theorem 11 needs to combine the result of Theorem 7 and the regret due to the estimation of the parameters similarly to what is done in Theorem 2.

Proof. *(Theorem 11)*

Step 1: payments and the regret. Similar to the proof of Theorem 7, we use the form of the VCG payments as in Eq. 24:

$$p_i^*(\mathbf{v}) = \Lambda_{\pi(i;f^*(\mathbf{v}))} q_i v_i - \int_0^{v_i} \Lambda_{\pi(i;f^*(\mathbf{v}_{-i},u))} q_i du,$$

while A–VCG3 uses the contingent payments in Eq. 28, which in expectation become

$$\tilde{p}_i'(\mathbf{v}) = \mathbb{E}_{\mathbf{x}}\big[\Lambda_{\pi(i;\tilde{f}(\mathbf{x}))}|\mathbf{v}\big] q_i v_i - \int_0^{v_i} \mathbb{E}_{\mathbf{x}}\big[\Lambda_{\pi(i;\tilde{f}(\mathbf{x}))}|\mathbf{v}_{-i}, u\big] q_i du. \tag{E.1}$$

We also need to introduce the expected payments

$$\tilde{p}_i(\mathbf{v}) = \Lambda_{\pi(i;\tilde{f}(\mathbf{v}))} q_i v_i - \int_0^{v_i} \Lambda_{\pi(i;\tilde{f}(\mathbf{v}_{-i},u))} q_i du,$$

which correspond to the VCG payments except for the use of allocation function $\tilde{f}$ in place of f^*.

Initially, we compute an upper bound over the per–ad regret $r_i = p_i^* - p_i'$ for each step of the exploitation phase and we later use this result to compute the upper bound for the regret R_T over T steps. We divide the per–ad regret in two different components:

$$\begin{aligned} r_i(\mathbf{v}) &= p_i^*(\mathbf{v}) - \tilde{p}_i'(\mathbf{v}) \\ &= \underbrace{p_i^*(\mathbf{v}) - p_i^{*\prime}(\mathbf{v})}_{\text{cSRP regret}} + \underbrace{p_i^{*\prime}(\mathbf{v}) - \tilde{p}_i'(\mathbf{v})}_{\text{learning regret}} = r_i^B(\mathbf{v}) + r_i^L(\mathbf{v}), \end{aligned} \tag{E.2}$$

where

- $r_i^B(\mathbf{v})$ is the regret due to the use of the approach proposed in [23] instead of the VCG payments, when all the parameters are known;
- $r_i^L(\mathbf{v})$ is the regret due to the uncertainty on the parameters when the payments defined in [23] are considered.

For the definitions of $\mathbf{s}$ and $\mathbb{E}_{\mathbf{x}|\mathbf{s}}[\Lambda_{\pi(i;f(\mathbf{x}))}|\mathbf{v}]$ refer to the proof of Theorem 7.
Step 2: the per–ad per–step cSRP regret. We can reuse the result obtained in the proof of Theorem 7. In particular, we can use the bound in Eq. D.2, i.e. $r_i^B(\mathbf{v}) \leq (K+1)\,\mu v_{\max}$. Given that we have assumed $N > K$, in the remaining parts of this proof we will use the following upper bound: $r_i^B(\mathbf{v}) \leq (K+1)\,\mu v_{\max} \leq N\mu v_{\max}$.

Step 3: the per–ad per–step learning regret. Similar to the previous step, we write the learning expected payments based on the cSRP in Eq. E.1 as

$$\tilde{p}'_i(\mathbf{v}) = \mathbb{P}[\mathbf{s}=\mathbf{1}]\tilde{p}_i(\mathbf{v}) + \\ + \mathbb{P}[\mathbf{s}\neq\mathbf{1}]\left(\mathbb{E}_{\mathbf{x}|\mathbf{s}\neq\mathbf{1}}[\Lambda_{\pi(i;\tilde{f}(\mathbf{x}))}|\mathbf{v}]q_i v_i - \int_0^{v_i}\mathbb{E}_{\mathbf{x}|\mathbf{s}\neq\mathbf{1}}[\Lambda_{\pi(i;\tilde{f}(\mathbf{x}))}|\mathbf{v}_{-i},u]q_i du\right).$$

Then the per–ad regret is

$$\begin{aligned} r_i^L(\mathbf{v}) &= p_i^{*\prime}(\mathbf{v}) - \tilde{p}'_i(\mathbf{v}) \\ &= \mathbb{P}[\mathbf{s}=\mathbf{1}]\left(p_i^*(\mathbf{v}) - \tilde{p}_i(\mathbf{v})\right) + \\ &\quad + \mathbb{P}[\mathbf{s}\neq\mathbf{1}]\Bigg(\underbrace{\mathbb{E}_{\mathbf{x}|\mathbf{s}\neq\mathbf{1}}[\Lambda_{\pi(i;f^*(\mathbf{x}))}|\mathbf{v}]q_i v_i - \int_0^{v_i}\mathbb{E}_{\mathbf{x}|\mathbf{s}\neq\mathbf{1}}[\Lambda_{\pi(i;f^*(\mathbf{x}))}|\mathbf{v}_{-i},u]q_i du}_{\leq v_{\max}} + \\ &\qquad \underbrace{-\mathbb{E}_{\mathbf{x}|\mathbf{s}\neq\mathbf{1}}[\Lambda_{\pi(i;\tilde{f}(\mathbf{x}))}|\mathbf{v}]q_i v_i + \int_0^{v_i}\mathbb{E}_{\mathbf{x}|\mathbf{s}\neq\mathbf{1}}[\Lambda_{\pi(i;\tilde{f}(\mathbf{x}))}|\mathbf{v}_{-i},u]q_i du}_{=-r_{i,1}^B\leq 0}\Bigg) \\ &\leq p_i^*(\mathbf{v}) - \tilde{p}_i(\mathbf{v}) + \mathbb{P}\big[\exists j : s_j = 0\big]v_{\max} \\ &\leq p_i^*(\mathbf{v}) - \tilde{p}_i(\mathbf{v}) + \sum_{j\in\mathcal{N}}\mathbb{P}\big[s_j=0\big]v_{\max} = p_i^*(\mathbf{v}) - \tilde{p}_i(\mathbf{v}) + N\mu v_{\max}. \end{aligned}$$

We now simply notice that payments $\tilde{p}_i$ are WVCG payments corresponding to the estimated allocation function $\tilde{f}$ and can be written as

$$\tilde{p}_i(\mathbf{v}) = \frac{q_i}{\tilde{q}_i^+}\Big[\widetilde{SW}\big(\tilde{f}_{-i}(\mathbf{v}),\mathbf{v}\big) - \widetilde{SW}_{-i}\big(\tilde{f}(\mathbf{v}),\mathbf{v}\big)\Big],$$

which allows us to use the results stated in proof of Theorem 2 and from Eq. C.5 we can conclude that

$$\sum_{i:\pi(i;f^*(\mathbf{v})\leq K)}\left(p_i^*(\mathbf{v}) - \tilde{p}_i(\mathbf{v})\right) \leq 2v_{\max}\eta\left(\sum_{m=1}^{K}\Lambda_m\right) \leq 2Kv_{\max}\eta.$$

Step 4: cumulative regret. We now bring together the two per–step regrets and we have that at each step of the the exploitation phase we have the regret $r = \sum_{i=1}^{N} r_i$. We first notice that the expected per–step regret r_i

for each ad a_i is defined as the difference between the VCG payments $p_i^*(\mathbf{v})$ and the (expected) payments $p_i'(\mathbf{v})$ computed by the randomized mechanism when the estimates $\tilde{q}^+$ are used. We notice that $p_i^*(\mathbf{v})$ can be strictly positive only for the K displayed ads, while $p_i'(\mathbf{v}) \geq 0 \; \forall i \in \mathcal{N}$, due to the mechanism randomization. Thus, $p_i^*(\mathbf{v}) - p_i'(\mathbf{v}) > 0$ only for at most K ads. Thus we obtain the per–step regret

$$\begin{aligned} r &\leq \sum_{i:\pi(i;f^*(\mathbf{v}))\leq K} r_i(\mathbf{v}) = \sum_{i:\pi(i;f^*(\mathbf{v}))\leq K} \left(r_i^B(\mathbf{v}) + r_i^L(\mathbf{v})\right) \\ &\leq KN\mu v_{\max} + \sum_{i:\pi(i;f^*(\mathbf{v}))\leq K} \left(p_i^*\left(\mathbf{v}\right) - \tilde{p}_i\left(\mathbf{v}\right) + N\mu v_{\max}\right) \\ &\leq KN\mu v_{\max} + 2Kv_{\max}\eta + KN\mu v_{\max} = 2Kv_{\max}\eta + 2KN\mu v_{\max}. \end{aligned}$$

Finally, the global regret becomes

$$R_T \leq v_{\max} K \left[(T-\tau) \left(2\sqrt{\frac{N}{\tau} \log \frac{2N}{\delta}} + 2\mu N \right) + \tau + \delta T \right].$$

Step 5: parameters optimization. We first simplify further the previous bound as

$$R_T \leq v_{\max} K \left[T \left(2\sqrt{\frac{N}{\tau} \log \frac{2N}{\delta}} + 2\mu N \right) + \tau + \delta T \right]. \tag{E.3}$$

We first optimize the value of τ, take the derivative of the previous bound w.r.t. τ and set it to zero and obtain

$$v_{\max} K \Big(-\tau^{-\frac{3}{2}} T \sqrt{N \log \frac{2N}{\delta}} + 1 \Big) = 0,$$

which leads to

$$\tau = T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log \frac{2N}{\delta} \right)^{\frac{1}{3}}.$$

Once replaced into Eq. E.3 we obtain

$$R_T \leq v_{\max} K \left[3T^{\frac{2}{3}} N^{\frac{1}{3}} \Big(\log \frac{2N}{\delta} \Big)^{\frac{1}{3}} + 2T\mu N + \delta T \right].$$

The optimization of the asymptotic order of the bound can then be obtained by setting μ and δ so as to equalize the orders of the second and third term in the bound. In particular, by setting

$$\mu = T^{-\frac{1}{3}} N^{-\frac{2}{3}} \quad \text{and} \quad \delta = T^{-\frac{1}{3}} N^{\frac{1}{3}},$$

we obtain the final bound

$$R_T \leq 6 v_{\max} K T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log \left(2 N^{\frac{2}{3}} T^{\frac{1}{3}} \right) \right)^{\frac{1}{3}}.$$

Notice that, since $\delta < 1$, this implies that $T > N$, and, since $\mu < 1$, it must be the case $T > N^{-2} > 1$ that always holds. Moreover $T > \tau$, thus $T > N \log \frac{2N}{\delta}$.

Appendix F. Proof of Revenue Regret in Theorem 14

Before deriving the proof of Theorem 14, we prove two lemmas that we use in what follows.

Lemma 2. *Let $\mathcal{G}$ be an arbitrary space of allocation functions, then for any $g \in \mathcal{G}$, when $|q_i - \tilde{q}_i^+| \leq \eta$ with probability $1 - \delta$, then for any $\hat{\mathbf{v}}$ we have*

$$-2K v_{\max} \eta \leq SW(g(\hat{\mathbf{v}}), \hat{\mathbf{v}}) - \widetilde{SW}(g(\hat{\mathbf{v}}), \hat{\mathbf{v}}) \frac{q_i}{\tilde{q}_i^+} \leq \frac{2K v_{\max}}{q_{\min}} \eta,$$

with probability $1 - \delta$.

Proof. By using the definition of SW and $\widetilde{\text{SW}}$ we have the following sequence of inequalities

$$\begin{aligned}
\widetilde{\text{SW}}(g(\hat{\mathbf{v}}), \hat{\mathbf{v}}) \frac{q_i}{\tilde{q}_i^+} - & \text{SW}(g(\hat{\mathbf{v}}), \hat{\mathbf{v}}) \\
&= \sum_{j : \pi(j; g(\hat{\mathbf{v}})) \leq K} \Gamma_{\pi(j; g(\hat{\mathbf{v}}))}(g(\hat{\mathbf{v}})) \hat{v}_j \left(\tilde{q}_j^+ \frac{q_i}{\tilde{q}_i^+} - q_j \right) \\
&\leq v_{\max} \sum_{j : \pi(j; g(\hat{\mathbf{v}})) \leq K} \left(\tilde{q}_j^+ \frac{q_i}{\tilde{q}_i^+} - q_j \right) \\
&\leq v_{\max} \sum_{j : \pi(j; g(\hat{\mathbf{v}})) \leq K} (\tilde{q}_j^+ - q_j) \leq 2K v_{\max} \eta.
\end{aligned}$$

The second statement follows from

$$\begin{aligned}
&\mathrm{SW}(g(\hat{\mathbf{v}}),\hat{\mathbf{v}})-\widetilde{\mathrm{SW}}(g(\hat{\mathbf{v}}),\hat{\mathbf{v}})\frac{q_i}{\tilde{q}_i^+} \\
&\qquad\leq \sum_{j:\pi(j;g(\hat{\mathbf{v}}))\leq K}\Gamma_{\pi(j;g(\hat{\mathbf{v}}))}\hat{v}_j\left(q_j-\tilde{q}_j^+\frac{q_i}{\tilde{q}_i^+}\right) \\
&\qquad\leq v_{\max}\sum_{j:\pi(j;g(\hat{\mathbf{v}}))\leq K}\left(q_j-q_j\frac{q_i}{\tilde{q}_i^+}+q_j\frac{q_i}{\tilde{q}_i^+}-\tilde{q}_j^+\frac{q_i}{\tilde{q}_i^+}\right) \\
&\qquad= v_{\max}\sum_{j:\pi(j;g(\hat{\mathbf{v}}))\leq K}\left[q_j\left(\frac{\tilde{q}_i^+-q_i}{\tilde{q}_i^+}\right)+\underbrace{(q_j-\tilde{q}_j^+)}_{\leq 0}\frac{q_i}{\tilde{q}_i^+}\right] \\
&\qquad\leq \frac{v_{\max}}{q_{\min}}\sum_{j:\pi(j;g(\hat{\mathbf{v}}))\leq K}(\tilde{q}_i-q_i+\eta)\leq\frac{2Kv_{\max}}{q_{\min}}\eta.
\end{aligned}$$

□

Lemma 3. *Let $\mathcal{G}$ be an arbitrary space of allocation functions, then for any $g\in\mathcal{G}$, when $|q_i-\tilde{q}_i^+|\leq\eta$ with probability $1-\delta$, we have*

$$0\leq\left(\widetilde{SW}(g(\hat{\mathbf{v}}),\hat{\mathbf{v}})-SW(g(\hat{\mathbf{v}}),\hat{\mathbf{v}})\right)\leq 2Kv_{\max}\eta,$$

with probability $1-\delta$.

Proof. The first inequality follows from

$$\begin{aligned}
&\mathrm{SW}(g(\hat{\mathbf{v}}),\hat{\mathbf{v}})-\widetilde{\mathrm{SW}}(g(\hat{\mathbf{v}}),\hat{\mathbf{v}}) \\
&\qquad=\sum_{j:\pi(j;g(\hat{\mathbf{v}}))\leq K}\Gamma_{\pi(j;g(\hat{\mathbf{v}}))}(g(\hat{\mathbf{v}}))\hat{v}_j\left(q_j-\tilde{q}_j^+\right) \\
&\qquad\leq v_{\max}\sum_{j:\pi(j;g(\hat{\mathbf{v}}))\leq K}(q_j-\tilde{q}_j^+)\leq 0,
\end{aligned}$$

while the second inequality follows from

$$\begin{aligned}
&\widetilde{\mathrm{SW}}(g(\hat{\mathbf{v}}),\hat{\mathbf{v}})-\mathrm{SW}(g(\hat{\mathbf{v}}),\hat{\mathbf{v}}) \\
&\qquad=\sum_{j:\pi(j;g(\hat{\mathbf{v}}))\leq K}\Gamma_{\pi(j;g(\hat{\mathbf{v}}))}(g(\hat{\mathbf{v}}))\hat{v}_j\left(\tilde{q}_j^+-q_j\right)
\end{aligned}$$

$$\leq v_{\max} \sum_{j:\pi(j;g(\hat{\mathbf{v}}))\leq K} \left(\tilde{q}_j^+ - q_j\right)$$
$$= v_{\max} \sum_{j:\pi(j;g(\hat{\mathbf{v}}))\leq K} (\tilde{q}_j + \eta - q_j) \leq 2K v_{\max}\eta.$$

□

We are now ready to proceed with the proof of Theorem 14.

PROOF. *(Theorem 14)*
Step 1: per–ad per–step regret. We first compute the per–step per–ad regret $r_i = p_i^*(\mathbf{v}) - \tilde{p}_i(\mathbf{v})$ at each step of the exploitation phase for each ad a_i. According to the definition of payments we have

$$r_i = \underbrace{\mathrm{SW}(f_{-i}^*(\mathbf{v}),\mathbf{v}) - \widetilde{\mathrm{SW}}(\tilde{f}_{-i}(\mathbf{v}),\mathbf{v})\frac{q_i}{\tilde{q}_i^+}}_{r_i^1} + \underbrace{\widetilde{\mathrm{SW}}_{-i}(\tilde{f}(\mathbf{v}),\mathbf{v})\frac{q_i}{\tilde{q}_i^+} - \mathrm{SW}_{-i}(f^*(\mathbf{v}),\mathbf{v})}_{r_i^2}.$$

We bound the first term through Lemma 2 and the following inequalities

$$r_i^1 = \mathrm{SW}(f_{-i}^*(\mathbf{v}),\mathbf{v}) - \widetilde{\mathrm{SW}}(f_{-i}^*(\mathbf{v}),\mathbf{v})\frac{q_i}{\tilde{q}_i^+} + \widetilde{\mathrm{SW}}(f_{-i}^*(\mathbf{v}),\mathbf{v})\frac{q_i}{\tilde{q}_i^+} - \widetilde{\mathrm{SW}}(\tilde{f}_{-i}(\mathbf{v}),\mathbf{v})\frac{q_i}{\tilde{q}_i^+}$$
$$\leq \max_{f\in\mathcal{F}_{-i}}\left(\mathrm{SW}(f(\mathbf{v}),\mathbf{v}) - \widetilde{\mathrm{SW}}(f(\mathbf{v}),\mathbf{v})\frac{q_i}{\tilde{q}_i^+}\right) + \underbrace{\left(\widetilde{\mathrm{SW}}(f_{-i}^*(\mathbf{v}),\mathbf{v}) - \max_{f\in\mathcal{F}_{-i}}\widetilde{\mathrm{SW}}(\tilde{f}(\mathbf{v}),\mathbf{v})\right)}_{\leq 0}\frac{q_i}{\tilde{q}_i^+}$$
$$\leq \frac{2Kv_{\max}}{q_{\min}}\eta,$$

with probability $1-\delta$. We rewrite r_i^2 as

$$r_i^2 = \left(\widetilde{\mathrm{SW}}(\tilde{f}(\mathbf{v}),\mathbf{v}) - \Gamma_{\pi(i;\tilde{f}(\mathbf{v}))}(\tilde{f}(\mathbf{v}))\tilde{q}_i^+ v_i\right)\frac{q_i}{\tilde{q}_i^+} - \mathrm{SW}(f^*(\mathbf{v}),\mathbf{v}) + \Gamma_{\pi(i;f^*(\mathbf{v}))}(f^*(\mathbf{v}))q_i v_i$$
$$= \underbrace{\widetilde{\mathrm{SW}}(\tilde{f}(\mathbf{v}),\mathbf{v})\frac{q_i}{\tilde{q}_i^+} - \mathrm{SW}(f^*(\mathbf{v}),\mathbf{v})}_{r_i^3} + \left(\Gamma_{\pi(i;f^*(\mathbf{v}))}\left(f^*(\mathbf{v})\right) - \Gamma_{\pi(i;\tilde{f}(\mathbf{v}))}(\hat{f}(\mathbf{v}))\right)q_i v_i.$$

We now focus on the term r_i^3 and use Lemma 2 to bound it as

$$r_i^3 = \widetilde{\mathrm{SW}}(\tilde{f}(\mathbf{v}),\mathbf{v})\frac{q_i}{\tilde{q}_i^+} - \mathrm{SW}(\tilde{f}(\mathbf{v}),\mathbf{v}) + \underbrace{\mathrm{SW}(\tilde{f}(\mathbf{v}),\mathbf{v}) - \max_{f\in\mathcal{F}}\mathrm{SW}(f(\mathbf{v}),\mathbf{v})}_{\leq 0}$$

$$\leq \max_{f\in\mathcal{F}} \left(\widetilde{\mathrm{SW}}(f(\mathbf{v}),\mathbf{v}) \frac{q_i}{\tilde{q}_i^+} - \mathrm{SW}(f(\mathbf{v}),\mathbf{v}) \right)$$
$$\leq 2K v_{\max} \eta.$$

Step 2: exploitation and cumulative regret. We define $I = \{i \in \mathcal{N} |\ \pi(i; \tilde{f}(\mathbf{v})) \leq K \vee \pi(i; \tilde{f}(\mathbf{v})) \leq K\}$, $|I| \leq 2K$. It is clear that only the ads a_i s.t. $i \in I$ have a regret $r_i \neq 0$. The other ads, $i \notin I$, have both $p_i^*(\mathbf{v}) = 0$ and $\tilde{p}_i(\mathbf{v}) = 0$. Thus, we can bound the regret r, at each exploitative step, in the following way

$$\begin{aligned}
r &= \sum_{i\in I} (r_i^1 + r_i^2) \\
&\leq \sum_{i\in I} \left(\frac{2K v_{\max}}{q_{\min}} \eta + 2K v_{\max} \eta \right) + \sum_{i\in I} \left(\Gamma_{\pi(i;f^*(\mathbf{v}))}(f^*(\mathbf{v})) - \Gamma_{\pi(i;\tilde{f}(\mathbf{v}))}(\tilde{f}(\mathbf{v})) \right) q_i v_i \\
&= \sum_{i\in I} \left(\frac{2K v_{\max}}{q_{\min}} \eta + 2K v_{\max} \eta \right) + \sum_{i=1}^{N} \left(\Gamma_{\pi(i;f^*(\mathbf{v}))}(f^*(\mathbf{v})) - \Gamma_{\pi(i;\tilde{f}(\mathbf{v}))}(\tilde{f}(\mathbf{v})) \right) q_i v_i \\
&\leq \frac{8K^2 v_{\max}}{q_{min}} \eta + \mathrm{SW}(f^*(\mathbf{v}),\mathbf{v}) - \mathrm{SW}(\tilde{f}(\mathbf{v}),\mathbf{v}) \\
&= \frac{8K^2 v_{\max}}{q_{min}} \eta + \mathrm{SW}(f^*(\mathbf{v}),\mathbf{v}) - \widetilde{\mathrm{SW}}(f^*(\mathbf{v}),\mathbf{v}) + \\
&\quad + \underbrace{\widetilde{\mathrm{SW}}(f^*(\mathbf{v}),\mathbf{v}) - \max_{f\in\mathcal{F}} \widetilde{\mathrm{SW}}(f)}_{\leq 0} + \widetilde{\mathrm{SW}}(\tilde{f}(\mathbf{v}),\mathbf{v}) - \mathrm{SW}(\tilde{f}(\mathbf{v}),\mathbf{v}) \\
&\leq \frac{8K^2 v_{\max}}{q_{min}} \eta + \underbrace{\mathrm{SW}(f^*(\mathbf{v}),\mathbf{v}) - \widetilde{\mathrm{SW}}(f^*(\mathbf{v}),\mathbf{v})}_{r^1} + \underbrace{\widetilde{\mathrm{SW}}(\tilde{f}(\mathbf{v}),\mathbf{v}) - \mathrm{SW}(\tilde{f}(\mathbf{v}),\mathbf{v})}_{r^2}.
\end{aligned}$$

The remaining terms r^1 and r^2 can be easily bounded using Lemma 3 as

$$r^1 \leq 0 \quad \text{and} \quad r^2 \leq 2K v_{\max} \eta.$$

Summing up all the terms we finally obtain

$$r \leq \frac{10K^2 v_{\max}}{q_{\min}} \eta$$

with probability $1-\delta$. Now, considering the per–step regret of the exploration and exploitation phases, we obtain the final bound on the cumulative regret

R_T as follows

$$R_T \leq v_{\max} K \left[\frac{10K}{\Gamma_{\min} q_{\min}} (T - \tau) \sqrt{\frac{N}{K\tau} \log \frac{2N}{\delta}} + \tau + \delta T \right].$$

Step 3: parameter optimization. Let $c := \frac{5}{\Gamma_{\min} q_{\min}}$, then we first simplify the previous bound as

$$R_T \leq v_{\max} K \left[2cT \sqrt{\frac{NK}{\tau} \log \frac{2N}{\delta}} + \tau + \delta T \right].$$

Taking the derivative w.r.t. τ leads to

$$v_{\max} K \Big(- \tau^{-\frac{3}{2}} cT \sqrt{NK \log \frac{2N}{\delta}} + 1 \Big) = 0,$$

which leads to

$$\tau = c^{\frac{2}{3}} K^{\frac{1}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log \frac{2N}{\delta} \right)^{\frac{1}{3}}.$$

Once replaced in the bound, we obtain

$$R_T \leq v_{\max} K \left[3c^{\frac{2}{3}} K^{\frac{1}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \Big(\log \frac{2N}{\delta} \Big)^{\frac{1}{3}} + \delta T \right].$$

Finally, we choose δ to optimize the asymptotic order by setting

$$\delta = c^{\frac{2}{3}} K^{\frac{1}{3}} T^{-\frac{1}{3}} N^{\frac{1}{3}},$$

which leads to the final bound

$$R_T \leq 4 v_{\max} c^{\frac{2}{3}} K^{\frac{4}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log \frac{2N^{\frac{2}{3}} T^{\frac{1}{3}}}{K^{\frac{1}{3}} c^{\frac{2}{3}}} \right)^{\frac{1}{3}}.$$

Notice that this bound imposes constraints on the value of T, indeed, $T > \tau$, thus $T > c^{\frac{2}{3}} K^{\frac{1}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log \frac{2N}{\delta} \right)^{\frac{1}{3}}$ and $\delta < 1$, thus $T > c^2 KN$, leading to:

$$T > c^2 KN \max \left\{ \log \frac{2N}{\delta}, 1 \right\}.$$

The problem associated with the previous bound is that τ and δ depends on $q_{\min}$, which is an unknown quantity. Thus actually choosing this value to optimize the bound may be unfeasible. An alternative choice of τ and δ is obtained by optimizing the bound removing the dependency on $q_{\min}$. Let $d := \frac{5}{\Gamma_{\min}}$, then we choose

$$\tau = d^{\frac{2}{3}} K^{\frac{1}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log \frac{2N}{\delta} \right)^{\frac{1}{3}},$$

and

$$\delta = K^{\frac{1}{3}} N^{\frac{1}{3}} d^{\frac{2}{3}} T^{-\frac{1}{3}}.$$

This leads to the final bound

$$R_T \leq 4 v_{\max} K^{\frac{4}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \frac{d^{\frac{2}{3}}}{q_{\min}} \left(\log \frac{2N^{\frac{2}{3}} T^{\frac{1}{3}}}{K^{\frac{1}{3}} d^{\frac{2}{3}}} \right)^{\frac{1}{3}}$$

Given that $\delta < 1$, this implies that $T > KNd^2$, and $T > \tau$ implies $T > d^2 KN \log \frac{2N}{\delta}$. Together they impose

$$T > d^2 KN \max \left\{ \log \frac{2N}{\delta}, 1 \right\}.$$

□

Appendix G. Deviation Regret

The definition of regret in Eq. 5 measures the cumulative difference between the revenue of the VCG compared to the one obtained by A–VCG1 over T steps. Upper–bounds over this quantity guarantees that the loss in terms of revenue does not increase linearly with T. As illustrated in the previous sections, the key passage in the proofs is the upper–bounding of the regret at each step of the exploitation phase (i.e., $r = \sum_{i=1}^{N} (p_i^*(\mathbf{v}) - \tilde{p}_i(\mathbf{v}))$). Nonetheless, we notice that this quantity could be negative as well. In this section we introduce a different notion of regret ($\tilde{R}_T$) that we study only for A–VCG1, leaving for the future a more comprehensive analysis of the other algorithms. Let us consider the following simple example. Let $N = 3$, $K = 1$, $v_i = 1$ for all the ads, and $q_1 = 0.1$, $q_2 = 0.2$, and $q_3 = 0.3$. Let assume that after the exploration phase we have $\tilde{q}_1^+ = 0.1$, $\tilde{q}_2^+ = 0.29$, $\tilde{q}_3^+ = 0.3$. The standard

VCG mechanism allocates ad a_3 and asks for a payment $p_3^*(\mathbf{v}) = 0.2$. During the exploitation phase A–VCG1 also allocates a_3 but asks for an (expected) payment $\tilde{p}_3(\mathbf{v}) = (\tilde{q}_2^+/\tilde{q}_3^+)q_3 = 0.29$. Thus, the regret in each exploitation step is $r = p_3^*(\mathbf{v}) - \tilde{p}_3(\mathbf{v}) = -0.09$. Although this result might seem surprising, it is due to the fact that while both A–VCG1 and VCG are truthful, in general A–VCG1 is not AE. We recall that a mechanism is AE if for any set of advertisers it always maximizes the social welfare. In the example, if for instance the true quality of ad a_3 is $q_3 = 0.28$, then the allocation induced by $\tilde{q}^+$s is not efficient anymore. For this reason, we characterized the behavior of A–VCG1 compared to the VCG considering the *deviation* between their payments, defined as

$$\tilde{R}_T(\mathfrak{A}) = \sum_{t=1}^{T} \Big| \sum_{i=1}^{N} (p_{i,t}^* - \tilde{p}_{i,t}) \Big|. \tag{G.1}$$

where we consider the definition of $\tilde{p}_i$ in Eq. C.2. We leave to further investigation the study of this regret for the other considered mechanisms.

Theorem 18. *Let us consider a sequential auction with N advertisers, K slots, and T steps with position–dependent cascade model with parameters $\{\Lambda_m\}_{m=1}^K$ and accuracy η as defined in Eq. 14. For any parameter $\tau \in \{1, \dots, T\}$ and $\delta \in (0, 1)$, the A–VCG1 achieves an auctioneer's revenue expected regret:*

$$\tilde{R}_T \leq K v_{\max} \left(\frac{2}{q_{\min}} (T - \tau)\, \eta + \tau + \delta T \right) \tag{G.2}$$

where $q_{\min} = \min_{i \in \mathcal{N}} q_i$. By setting the parameters to

- $\delta = N^{\frac{1}{3}} K^{-\frac{1}{3}} T^{-\frac{1}{3}}$,
- $\tau = \frac{K^{-\frac{1}{3}} N^{\frac{1}{3}} T^{\frac{2}{3}}}{\Lambda_{\min}^{\frac{2}{3}}} \left(\log \frac{N}{\delta} \right)^{\frac{1}{3}}$,

the regret is

$$\tilde{R}_T \leq 4 \frac{K^{-\frac{1}{3}} N^{\frac{1}{3}} T^{\frac{2}{3}}}{q_{\min} \Lambda_{\min}^{\frac{2}{3}}} \left(\log N^{\frac{2}{3}} K^{\frac{1}{3}} T^{\frac{1}{3}} \right)^{\frac{1}{3}}. \tag{G.3}$$

PROOF. We initially provide a bound over the per–step regret during the exploitation phase. We consider the two sides of the bound separately. We have that for the first side of the bound we can use the result provided in Step 3 in the proof of Theorem 2, i.e.,

$$\begin{aligned} r_1 &= \sum_{m=1}^{K} (p^*_{\alpha(m;\theta^*)}(\mathbf{v}) - \tilde{p}_{\alpha(m;\tilde{\theta})}(\mathbf{v})) \\ &\leq 2K v_{\max} \eta, \end{aligned}$$

with probability $1-\delta$. Now we bound the other side.

$$\begin{aligned} r_2 &= \sum_{m=1}^{K} \left(\tilde{p}_{\alpha(m;\tilde{\theta})}(\mathbf{v}) - p^*_{\alpha(m;\theta^*)}(\mathbf{v}) \right) \\ &= \sum_{m=1}^{K} \sum_{l=m}^{K} \Delta_l \left(\frac{\max_{i \in \mathcal{N}} (\tilde{q}_i^+ v_i; l+1)}{\tilde{q}^+_{\alpha(m;\tilde{\theta})}} q_{\alpha(m;\tilde{\theta})} - \max_{i \in \mathcal{N}} (q_i v_i; l+1) \right) \\ &\leq \max_{i \in \mathcal{N}} (q_i v_i; l+1) \sum_{m=1}^{K} \sum_{l=m}^{K} \Delta_l \left(\frac{\max_{i \in \mathcal{N}} (\tilde{q}_i^+ v_i; l+1)}{\max_{i \in \mathcal{N}} (q_i v_i; l+1)} - 1 \right). \end{aligned}$$

In order to proceed we need to bound the ratio in the inner term. We first recall that for any ad a_i, we have that $\tilde{q}_i^+ v_i = (\tilde{q}_i + \eta)\, v_i \leq (q_i + 2\eta)\, v_i$. Let $i' \in \arg\max_{j \in \mathcal{N}} (q_j v_j; l+1)$ be the ad displayed in s_{l+1} when the true qualities are known. We distinguish two cases:

- The ad $a_{i'}$ shifts from slot $l+1$ to a higher precedence slot when allocated according to $\tilde{f}$, i.e., $\pi(i'; \tilde{f}(v)) \leq l+1$. In this case we have

$$\frac{\max_{i \in \mathcal{N}} (\tilde{q}_i^+ v_i; l+1)}{\max_{i \in \mathcal{N}} (q_i v_i; l+1)} \leq \frac{\tilde{q}^+_{i'} v_{i'}}{q_{i'} v_{i'}} \leq 1 + \frac{2\eta}{q_{\min}}.$$

- The ad $a_{i'}$ shifts from slot $l+1$ to a lower precedence slot when allocated according to $\tilde{f}$, i.e., $\pi(i'; \tilde{f}(v)) > l+1$. In this case, there must exist an ad j that in the exact allocation is allocated after i' but it is promoted to a higher precedence slot according to $\tilde{f}$. This corresponds to a $j \in \mathcal{N}$ such that $\pi(j; f^*(v)) \geq l+1$ but $\pi(j; \tilde{f}(v)) < l+1$. As a result we have

$$\frac{\max_{i \in \mathcal{N}} (\tilde{q}_i^+ v_i; l+1)}{\max_{i \in \mathcal{N}} (q_i v_i; l+1)} \leq \frac{\tilde{q}^+_j v_j}{q_{i'} v_{i'}} \leq \frac{q_j v_j + 2\eta v_j}{q_{i'} v_{i'}} \leq \frac{q_j v_j + 2\eta \frac{q_j}{q_{\min}} v_j}{q_{i'} v_{i'}}$$

$$\leq \frac{q_{i'}v_{i'} + 2\eta\frac{q_{i'}}{q_{\min}}v_{i'}}{q_{i'}v_{i'}} \leq 1 + \frac{2\eta}{q_{\min}}.$$

Using these results we obtain

$$\begin{aligned} r_2 &\leq v_{\max}\sum_{m=1}^{K}\sum_{l=m}^{K}\Delta_l\left(1 + \frac{1}{q_{\min}}2\eta - 1\right) \\ &\leq v_{\max}\frac{1}{q_{\min}}2\eta\sum_{m=1}^{K}\underbrace{\sum_{l=m}^{K}\Delta_l}_{=\Lambda_m} \leq v_{\max}\frac{1}{q_{\min}}2\eta K. \end{aligned}$$

with probability $1-\delta$. As a result we have

$$\left|\sum_{m=1}^{K}(p^*_{\alpha(m;\theta^*)}(\mathbf{v}) - \tilde{p}_{\alpha(m;\tilde{\theta})}(\mathbf{v}))\right| \leq 2v_{\max}K\frac{\eta}{q_{\min}},$$

with probability $1-\delta$. The final bound on the expected regret is thus

$$\tilde{R}_T \leq Kv_{\max}\left(\frac{2}{q_{\min}}(T-\tau)\,\eta + \tau + \delta T\right). \tag{G.4}$$

We first simplify the previous bound as

$$\begin{aligned} \tilde{R}_T &\leq Kv_{\max}\left(\frac{2T}{q_{\min}}\sqrt{\left(\sum_{m=1}^{K}\frac{1}{\Lambda_m^2}\right)\frac{N}{K^2\tau}\log\frac{N}{\delta}} + \tau + \delta T\right) \\ &\leq Kv_{\max}\left(\frac{2T}{q_{\min}\Lambda_{\min}}\sqrt{\frac{N}{K\tau}\log\frac{N}{\delta}} + \tau + \delta T\right) \end{aligned}$$

and choosing the parameters

$$\tau = \frac{K^{-\frac{1}{3}}N^{\frac{1}{3}}T^{\frac{2}{3}}}{\Lambda_{\min}^{\frac{2}{3}}}\left(\log\frac{N}{\delta}\right)^{\frac{1}{3}},$$

$$\delta = N^{\frac{1}{3}}K^{-\frac{1}{3}}T^{-\frac{1}{3}},$$

the final bound is

$$\tilde{R}_T \leq 4\frac{K^{-\frac{1}{3}}N^{\frac{1}{3}}T^{\frac{2}{3}}}{q_{\min}\Lambda_{\min}^{\frac{2}{3}}}\left(\log N^{\frac{2}{3}}K^{\frac{1}{3}}T^{\frac{1}{3}}\right)^{\frac{1}{3}}.$$

The fact $\delta < 1$ implies that $T > \frac{N}{K}$, and $T > \tau$ implies $T > \frac{K^{-1}N}{\Lambda_{\min}^2}\log\frac{N}{\delta}$. Together they constrain

$$T > \frac{N}{K\Lambda_{\min}^2}\max\left\{\log\frac{N}{\delta}, 1\right\}.$$

□

Remark (the bound). We notice that the bound is very similar to the bound for the regret R_T but now an inverse dependency on $q_{\min}$ appears. This suggests that bounding the deviation between the two mechanisms is more difficult than bounding the revenue loss and that, as the qualities become smaller, the A–VCG1 could be less and less efficient and, thus, have a larger and larger revenue. This result has two important implications. *(i)* If SW maximization is an important requirement in the design of the learning mechanism, we should analyze the loss of A–VCG1 in terms of social welfare and provide (probabilistic) guarantees about the number of steps the learning mechanism need in order to be AE (see [17] for a similar analysis). *(ii)* If social welfare is not a priority, this result implies that a learning mechanism could be preferable w.r.t. the standard VCG mechanism. We believe that further theoretical analysis and experimental validation are needed to understand better both aspects.

Appendix H. Proofs of Social-Welfare Regret in Theorems 3 and 15

Before stating the main result of this section, we need the following technical lemma.

Lemma 4. *Let us consider an auction with N advertisers, K slots, and T steps, and a mechanism that separates the exploration (τ steps) and the exploitation phases ($T-\tau$ steps). Consider an arbitrary space of allocation functions $\mathcal{G}$, $\tilde{g} \in \arg\max_{g'\in\mathcal{G}} \widetilde{SW}(g'(\hat{\mathbf{v}}), \hat{\mathbf{v}})$ and $|q_i - \tilde{q}_i^+| \leq \eta$ with probability $1-\delta$. For any $g \in \mathcal{G}$, an upper bound of SW regret R_T^{SW} of the mechanism adopting $\tilde{g}$ instead of g is:*

$$R_T^{SW} \leq v_{\max}K\left[2(T-\tau)\eta + \tau + \delta T\right].$$

PROOF. We now prove the bound on the social welfare, starting from the cumulative per–step regret during the exploitation phase.

$$\begin{aligned} r &= \mathrm{SW}(g(\hat{\mathbf{v}}), \hat{\mathbf{v}}) - \mathrm{SW}(\tilde{g}(\hat{\mathbf{v}}), \hat{\mathbf{v}}) \\ &= \mathrm{SW}(g(\hat{\mathbf{v}}), \hat{\mathbf{v}}) - \widetilde{\mathrm{SW}}(g(\hat{\mathbf{v}}), \hat{\mathbf{v}}) + \\ &\quad + \underbrace{\widetilde{\mathrm{SW}}(g(\hat{\mathbf{v}}), \hat{\mathbf{v}}) - \max_{g' \in \mathcal{G}} \widetilde{\mathrm{SW}}(g'(\hat{\mathbf{v}}), \hat{\mathbf{v}})}_{\leq 0} + \widetilde{\mathrm{SW}}(\tilde{g}(\hat{\mathbf{v}}), \hat{\mathbf{v}}) - \mathrm{SW}(\tilde{g}(\hat{\mathbf{v}}), \hat{\mathbf{v}}) \\ &\leq \underbrace{\mathrm{SW}(g(\hat{\mathbf{v}}), \hat{\mathbf{v}}) - \widetilde{\mathrm{SW}}(g(\hat{\mathbf{v}}), \hat{\mathbf{v}})}_{r^1} + \underbrace{\widetilde{\mathrm{SW}}(\tilde{g}(\hat{\mathbf{v}}), \hat{\mathbf{v}}) - \mathrm{SW}(\tilde{g}(\hat{\mathbf{v}}), \hat{\mathbf{v}})}_{r^2}. \end{aligned}$$

The two remaining terms r^1 and r^2 can be easily bounded by using Lemma 3

$$r \leq r_1 + r_2 \leq 0 + 2K v_{\max} \eta = 2K v_{\max} \eta$$

with probability $1 - \delta$.

Thus, we can conclude that:

$$R_T^{SW} \leq v_{\max} K \left[2(T - \tau)\eta + \tau + \delta T \right].$$

□

PROOF. *(Theorem 3)*

Step 1: cumulative regret. We apply Lemma 4 to the position–dependent cascade model with $\{q_i\}_{i \in \mathcal{N}}$ unknowns, obtaining

$$\begin{aligned} R_T^{SW} &\leq v_{\max} K \left[2(T - \tau)\eta + \tau + \delta T \right] \\ &\leq v_{\max} K \left[\frac{2}{\Lambda_{\min}} (T - \tau) \sqrt{\frac{N}{K\tau} \log \frac{2N}{\delta}} + \tau + \delta T \right]. \end{aligned}$$

Step 2: parameter optimization. First we notice that adopting the value of the parameters identified in Theorem 2 we obtain an upper bound $\tilde{O}(T^{\frac{2}{3}})$ for the global regret R_T^{SW}.

In order to find values that better optimize the bound over R_T^{SW}, let $e := \frac{1}{\Lambda_{\min}}$, then we first simplify the previous bound as

$$R_T^{SW} \leq v_{\max} K \left[2e \sqrt{\frac{N}{K\tau} \log \frac{2N}{\delta}} + \tau + \delta T \right].$$

Taking the derivative of the previous bound w.r.t. τ leads to

$$v_{\max} K \left(-\tau^{-\frac{3}{2}} e T \sqrt{\frac{N}{K} \log \frac{2N}{\delta}} + 1 \right) = 0,$$

which leads to

$$\tau = e^{\frac{2}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} K^{-\frac{1}{3}} \left(\log \frac{2N}{\delta} \right)^{\frac{1}{3}}.$$

Once replaced in the bound, we obtain

$$R_T^{SW} \leq v_{\max} K \left[3 e^{\frac{2}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} K^{-\frac{1}{3}} \left(\log \frac{2N}{\delta} \right)^{\frac{1}{3}} + \delta T \right].$$

Finally, we choose δ to optimize the asymptotic order by setting

$$\delta = e^{\frac{2}{3}} K^{-\frac{1}{3}} T^{-\frac{1}{3}} N^{\frac{1}{3}}$$

The final bound is

$$R_T^{SW} \leq 4 v_{\max} e^{\frac{2}{3}} K^{\frac{2}{3}} T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log 2 e^{-\frac{2}{3}} N^{\frac{2}{3}} K^{\frac{1}{3}} T^{\frac{1}{3}} \right)^{\frac{1}{3}}.$$

Given that $\delta < 1$ this implies that $T > e^2 K^{-1} N$. This constrain is satisfied imposing $T > \tau$, i.e.,

$$T > e^2 K^{-1} N \log \frac{2N}{\delta}.$$

□

PROOF. *(Theorem 15)*
Step 1: cumulative regret. We apply Lemma 4 to the model with position– and ad–dependent externalities with $\{q_i\}_{i \in \mathcal{N}}$ unknowns, obtaining

$$\begin{aligned} R_T^{SW} &\leq v_{\max} K \left[2(T - \tau)\eta + \tau + \delta T \right] \\ &\leq v_{\max} K \left[\frac{2}{\Gamma_{\min}} (T - \tau) \sqrt{\frac{N}{K\tau} \log \frac{2N}{\delta}} + \tau + \delta T \right]. \end{aligned}$$

Step 2: parameter optimization. First we notice that adopting the value of the parameters identified in Theorem 14 we obtain an upper bound $\tilde{O}(T^{\frac{2}{3}})$ for the global regret R_T^{SW}.

In order to find values that better optimize the bound over R_T^{SW}, it is possible to use the procedure followed in the proof of Theorem 3 with $e := \frac{1}{\Gamma_{\min}}$:

$$R_T^{SW} \leq 4v_{\max}e^{\frac{2}{3}}K^{\frac{2}{3}}N^{\frac{1}{3}}T^{\frac{2}{3}}\left(\log 2e^{-\frac{2}{3}}N^{\frac{2}{3}}K^{\frac{1}{3}}T^{\frac{1}{3}}\right)^{\frac{1}{3}}.$$

□

Appendix I. Proof of Social-Welfare Regret in Theorem 8

PROOF. *(Theorem 8)*

The bound over the SW regret R_T^{SW} can be easily derived considering that each bid is modified by the cSRP with a probability of μ. Thus we can define $S' = \{\mathbf{s}'|\mathbf{s}' \in \{0,1\}^N, \pi(i; f^*(\hat{\mathbf{v}})) \leq K \Rightarrow s'_i = 1\}$, i.e., the set of the random realizations where the cSRP does not modify the bids of the ads displayed when the allocation function f^* is applied to the true bids $\hat{\mathbf{v}}$. Thus we have:

$$R_T^{SW} \leq T\left(\mathbb{P}\left[\mathbf{s} \in S'\right] \cdot 0 + \underbrace{\mathbb{P}\left[\mathbf{s} \notin S'\right]}_{\leq K\mu} K v_{\max}\right) \leq K^2 \mu v_{\max} T.$$

□

Appendix J. Proof of Social-Welfare Regret Theorem 12

PROOF. *(Theorem 12)*

Step 1: per–step regret. We start computing the per–step regret over the SW during the exploitation phase.

First of all we introduce the following definition: $S' = \{\mathbf{s}'|\mathbf{s}' \in \{0,1\}^N, \pi(i; f^*(\hat{\mathbf{v}})) \leq K \Rightarrow s'_i = 1\}$, i.e., the set of the random realization where the cSRP does not modify the bids of the ads displayed when the allocation function is f^* is applied to the true bids $\hat{\mathbf{v}}$.

We now provide the bound over the regret.

$$\begin{aligned} r &= \mathrm{SW}(f^*(\hat{\mathbf{v}}), \hat{\mathbf{v}}) - \mathbb{E}_{\mathbf{x}}\left[\mathrm{SW}(\tilde{f}(\mathbf{x}), \hat{\mathbf{v}})|\hat{\mathbf{v}}\right] \\ &= \underbrace{\mathbb{P}[\mathbf{s} \in S']}_{\leq 1}\left(\mathrm{SW}(f^*(\hat{\mathbf{v}}), \hat{\mathbf{v}}) - \mathbb{E}_{\mathbf{x}|\mathbf{s}\in S'}\left[\mathrm{SW}(\tilde{f}(x), \hat{\mathbf{v}})|\hat{\mathbf{v}}\right]\right) + \end{aligned}$$

$$
\begin{aligned}
&+ \underbrace{\mathbb{P}[\mathbf{s} \notin S']}_{\leq K\mu} \left(\mathrm{SW}(f^*(\hat{\mathbf{v}}), \hat{\mathbf{v}}) - \mathbb{E}_{\mathbf{x}|\mathbf{s}\notin S'} \left[\mathrm{SW}(\tilde{f}(x), \mathbf{v}) | \hat{\mathbf{v}} \right] \right) \\
&\leq \mathrm{SW}(f^*(\hat{\mathbf{v}}), \hat{\mathbf{v}}) - \mathbb{E}_{\mathbf{x}|\mathbf{s}\in S'} \left[\mathrm{SW}(\tilde{f}(\mathbf{x}), \hat{\mathbf{v}}) | \hat{\mathbf{v}} \right] + \\
&+ K\mu \underbrace{\left(\mathrm{SW}(f^*(\hat{\mathbf{v}}), \hat{\mathbf{v}}) - \underbrace{\mathbb{E}_{\mathbf{x}|\mathbf{s}\notin S'} \left[\mathrm{SW}(\tilde{f}(x), \mathbf{v}) | \hat{\mathbf{v}} \right]}_{\geq 0} \right)}_{\leq K v_{\max}} \\
&\leq \underbrace{\mathrm{SW}(f^*(\hat{\mathbf{v}}), \hat{\mathbf{v}}) - \mathbb{E}_{\mathbf{x}|\mathbf{s}\in S'} \left[\widetilde{\mathrm{SW}}(f^*(\mathbf{x}), \hat{\mathbf{v}}) | \hat{\mathbf{v}} \right]}_{r_1 \leq 0} + \\
&+ \underbrace{\mathbb{E}_{\mathbf{x}|\mathbf{s}\in S'} \left[\widetilde{\mathrm{SW}}(f^*(\mathbf{x}), \hat{\mathbf{v}}) | \hat{\mathbf{v}} \right] - \mathbb{E}_{\mathbf{x}|\mathbf{s}\in S'} \left[\widetilde{\mathrm{SW}}(\tilde{f}(\mathbf{x}), \hat{\mathbf{v}}) | \hat{\mathbf{v}} \right]}_{r_2 \leq 0} + \\
&+ \mathbb{E}_{\mathbf{x}|\mathbf{s}\in S'} \left[\widetilde{\mathrm{SW}}(\tilde{f}(\mathbf{x}), \hat{\mathbf{v}}) | \hat{\mathbf{v}} \right] - \mathbb{E}_{\mathbf{x}|\mathbf{s}\in S'} \left[\mathrm{SW}(\tilde{f}(\mathbf{x}), \hat{\mathbf{v}}) | \hat{\mathbf{v}} \right] + v_{\max} \mu K^2 \\
&\leq \max_{f \in \mathcal{F}} \left(\mathbb{E}_{\mathbf{x}|\mathbf{s}\in S'} \left[\widetilde{\mathrm{SW}}(f(\mathbf{x}), \hat{\mathbf{v}}) - \mathrm{SW}(f(\mathbf{x}), \hat{\mathbf{v}}) | \hat{\mathbf{v}} \right] \right) + v_{\max} \mu K^2 \\
&\leq \max_{f \in \mathcal{F}} \left(\sum_{j:\pi(j;f(x))\leq K} \Lambda_{\pi(j;f(x))} v_j (\tilde{q}_j - q_j) \right) + v_{\max} \mu K^2 \\
&\leq v_{\max} \max_{f \in \mathcal{F}} \left(\sum_{j:\pi(j;f(x))\leq K} (\tilde{q}_j - q_j) \right) + v_{\max} \mu K^2 \\
&\leq 2 v_{\max} K \eta + v_{\max} \mu K^2 = v_{\max} K \left(2\eta + K\mu \right).
\end{aligned}
$$

We provide a brief intuition of bounds r_1 and r_2. The bound r_1 can be explained noticing that when the bids of the ads displayed in $f^*(\hat{\mathbf{v}})$ are not modified we have that $\alpha(m; f^*(\hat{\mathbf{v}})) = \alpha(m; f^*(\mathbf{x}))$ where $m \leq K$ and $\mathbf{x}$ s.t. $\mathbf{s} \in S'$. The bound for r_2 can be understood noticing that when the bids of the ads s.t. $\pi(j; f^*(\mathbf{x})) \leq K$ are not modified and $x_i \leq \hat{v}_i \ \forall i \in \mathcal{N}$, we obtain $\widetilde{\mathrm{SW}}(f^*(\mathbf{x}), \hat{\mathbf{v}}) = \widetilde{\mathrm{SW}}(f^*(\mathbf{x}), \mathbf{x}) \leq \max_{\theta \in \Theta} \widetilde{\mathrm{SW}}(\theta, \mathbf{x}) = \widetilde{\mathrm{SW}}(\tilde{f}(\mathbf{x}), \mathbf{x}) \leq \widetilde{\mathrm{SW}}(\tilde{f}(\mathbf{x}), \hat{\mathbf{v}})$.

Step 2: cumulative regret. We can now compute the upper bound for the global regret

$$
R_T^{SW} \leq v_{\max} K \left[(T - \tau)(2\eta + K\mu) + \tau + \delta T \right]
$$

$$\leq v_{\max} K \left[(T-\tau) \left(2\sqrt{\frac{N}{\tau} \log \frac{2N}{\delta}} + K\mu \right) + \tau + \delta T \right].$$

Step 3: parameter optimization. We first simplify the previous bound as

$$R_T^{SW} \leq v_{\max} K \left[2T\sqrt{\frac{N}{\tau} \log \frac{2N}{\delta}} + K\mu T + \tau + \delta T \right].$$

Taking the derivative of the previous bound w.r.t. τ leads to

$$v_{\max} K \left(-\tau^{-\frac{3}{2}} T \sqrt{N \log \frac{2N}{\delta}} + 1 \right) = 0,$$

which leads to

$$\tau = T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log \frac{2N}{\delta} \right)^{\frac{1}{3}}.$$

Once replaced τ in the bound, we obtain

$$R_T^{SW} \leq 3 v_{\max} K T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log \frac{2N}{\delta} \right)^{\frac{1}{3}} + \mu K^2 v_{\max} T + \delta v_{\max} K T.$$

Finally, we choose δ and μ to optimize the asymptotic order by setting

$$\delta = T^{-\frac{1}{3}} N^{\frac{1}{3}},$$
$$\mu = K^{-1} T^{-\frac{1}{3}} N^{\frac{1}{3}}.$$

The final bound is

$$R_T^{SW} \leq 5 \cdot v_{\max} K T^{\frac{2}{3}} N^{\frac{1}{3}} \left(\log 2T^{\frac{1}{3}} N^{\frac{2}{3}} \right)^{\frac{1}{3}}.$$

Given that $\delta < 1$ this implies that $T > N$ and, given that $\mu < 1$ we have that $T > \frac{N}{K^3}$. Both the constraints are satisfied imposing $T > \tau$, i.e.,

$$T > N \log \frac{2N}{\delta}.$$

□